

Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data

Andreas Anastasiou

University of Oxford, Oxford, UK

E-mail: anastasi@stats.ox.ac.uk

August 2016

Abstract

The asymptotic normality of the maximum likelihood estimator (MLE) under regularity conditions is a cornerstone of statistical theory. In this paper, we give explicit upper bounds on the distributional distance between the distribution of the MLE of a vector parameter, and the multivariate normal distribution. We work with possibly high-dimensional independent but not necessarily identically distributed random vectors. In addition, we obtain explicit upper bounds even in cases where the MLE does not have a closed-form expression.

Key words : Multi-parameter maximum likelihood estimation; multivariate normal approximation; Stein's method

1 Introduction

In this paper, we give explicit upper bounds on the distributional distance between the distribution of a vector MLE and the multivariate normal, which under specific regularity conditions is the MLE's limiting distribution. We focus on independent but not necessarily identically distributed random vectors. The quantitative statement obtained from our bounds can be helpful to assess whether using the limiting distribution of the MLE is an acceptable approximation or not. From the opposite point of view, the results presented in this paper can save both money and time by giving a good indication on whether a larger sample size is indeed necessary, for a good approximation to hold. The wide applicability of the maximum likelihood estimation method adds to the importance of our results. Among others, an MLE is used in ordinary and generalised linear models, time series analysis and a large number of other situations related to hypothesis testing and confidence intervals. They appear in a broad category of different fields, such as econometrics, computational biology and data modelling in physics and psychology.

The notation which is used throughout the paper is as follows. The parameter space is $\Theta \subset \mathbb{R}^d$ equipped with the Euclidean norm. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ be a parameter from the parameter space, while $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}, \dots, \theta_{0,d})$ denotes the true, but unknown, value of the parameter. The probability density (or probability mass) function is denoted by $f(\mathbf{x}|\boldsymbol{\theta})$, where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. The likelihood function is $L(\boldsymbol{\theta}; \mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta})$. Its natural logarithm, called the log-likelihood function is denoted by $l(\boldsymbol{\theta}; \mathbf{x})$. A maximum likelihood estimate (not seen as a random vector) is a value of the parameter which maximises the likelihood function. For many models the maximum likelihood estimator as a random vector exists and is also unique, in which case it is denoted by $\hat{\boldsymbol{\theta}}_n(\mathbf{X})$; this is known as the 'regular' case. Existence and uniqueness of the MLE can not be taken for granted, see e.g. Billingsley (1961) for an example of non-uniqueness.

In order to secure existence and uniqueness in the case where the likelihood function $L(\boldsymbol{\theta}; \mathbf{x})$ is twice continuously differentiable varying in an open parameter space $\Theta \subset \mathbb{R}^d$, we make the following assumptions from [Makelainen et al. \(1981\)](#):

(A1) $\lim_{\boldsymbol{\theta} \rightarrow \partial\Theta} L(\boldsymbol{\theta}; \mathbf{x}) = 0$, where $\partial\Theta$ is the boundary of the parameter space,

(A2) the Hessian matrix

$$\mathbf{H}(\boldsymbol{\theta}; \mathbf{x}) = \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}; \mathbf{x}) \right\}_{i,j=1,\dots,d} \quad (\text{where } \frac{\partial}{\partial \theta_i} \text{ denotes partial derivatives})$$

of second partial derivatives is negative definite at every point $\boldsymbol{\theta} \in \Theta$ for which the gradient vector

$$\nabla L(\boldsymbol{\theta}; \mathbf{x}) = \left\{ \frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta}; \mathbf{x}) \right\}_{i=1,\dots,d}$$

vanishes.

The interest is on assessing the quality of the asymptotic normality of the MLE and the approach we follow is partly based on Stein's method under a multivariate setting. Let

$$H = \left\{ h : \mathbb{R}^d \rightarrow \mathbb{R} : h \text{ is three times differentiable with bounded derivatives} \right\} \quad (1)$$

be the class of test functions we use in this paper. We abbreviate $\|h\|_1 := \sup_i \left\| \frac{\partial}{\partial x_i} h \right\|$, $\|h\|_2 := \sup_{i,j} \left\| \frac{\partial^2}{\partial x_i \partial x_j} h \right\|$ and $\|h\|_3 := \sup_{i,j,k} \left\| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h \right\|$. With $Z \sim N(0, 1)$, our results give upper bounds on

$$\left| \mathbb{E} \left[h \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0 \right) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right|, \quad (2)$$

where $\bar{I}_n(\boldsymbol{\theta}_0)$ is defined in (3). The bounds are explicit in terms of the sample size and $\boldsymbol{\theta}_0$. The two main results of the paper are given in Theorems 2.2 and 3.1. Theorem 2.2 gives a general upper bound on (2) which holds under the usual, sufficient regularity conditions for the asymptotic normality of the MLE. The generality of the bound adds to its importance as it can be applied in various different occasions; we have chosen the class of linear regression models to serve as an illustration of our results. Theorem 3.1 is also substantial since, under further assumptions, we achieve to obtain upper bounds related to the asymptotic normality of the MLE, even when the MLE is not known analytically.

The paper is organised as follows. Section 2 first treats the case of independent but not necessarily identically distributed (i.n.i.d.) random vectors. The upper bound on the distributional distance between the distribution of the vector MLE and the multivariate normal distribution is presented. Special attention is given to linear regression models with an application to the simplest case of the straight-line model. Furthermore, under weaker regularity conditions, we explain how the bound can be simplified for the case of i.i.d. random vectors. Specific results for independent random variables that follow the normal distribution with unknown mean and variance are also given. Section 3 contains an upper bound on the aforementioned distributional distance, which holds even in cases where no analytic expression of the vector MLE is available. We illustrate the results through the Beta distribution with both shape parameters unknown. In order to make the paper more easily readable, we only provide an outline of the proofs of our main Theorems 2.2 and 3.1 and the complete proofs are given in Section 4. In addition, some technical results and proofs of corollaries that are not essential for the smooth understanding of the paper's developments are confined in the Appendix.

2 Bounds for multi-parameter distributions

In this section we examine the case of i.n.i.d. t -dimensional random vectors, for $t \in \mathbb{Z}^+$. Apart from the assumptions (A1) and (A2) for the existence and uniqueness of the MLE, we use some regularity conditions, first stated in [Hoadley \(1971\)](#), in order to establish the asymptotic normality of the MLE. We give an upper bound on the distributional distance between the distribution of the MLE and the multivariate normal and then we focus on the specific case of linear models. The last subsection covers, under weaker regularity conditions, the case of i.i.d. random vectors and an example from the normal distribution with unknown mean and variance serves as an illustration of our results.

2.1 A general bound

For $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ being i.n.i.d. random vectors, we denote by $f_i(\mathbf{x}, \boldsymbol{\theta})$ the probability density (or mass) function of \mathbf{X}_i . The likelihood function is $L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n f_i(\mathbf{x}_i | \boldsymbol{\theta})$, with its logarithm being denoted by $l(\boldsymbol{\theta}; \mathbf{x})$. Assuming that the parameter space Θ is an open subset of \mathbb{R}^d , (A1) and (A2) are sufficient for the existence and uniqueness of $\hat{\boldsymbol{\theta}}_n(\mathbf{X})$. To simplify notation let $\mathbb{E}[\cdot]$ and $\mathbb{P}[\cdot]$ denote expectation and probability with respect to $\boldsymbol{\theta}_0$. We work under the following regularity conditions for the asymptotic normality of the MLE to hold, ([Hoadley, 1971](#)):

- (N1) $\hat{\boldsymbol{\theta}}_n(\mathbf{X}) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$, as $n \rightarrow \infty$, where $\boldsymbol{\theta}_0$ is the true parameter value;
- (N2) The Hessian matrix $J_k(\mathbf{X}_k, \boldsymbol{\theta}) = \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f_k(\mathbf{X}_k | \boldsymbol{\theta})) \right\}_{i,j=1,2,\dots,d} \in \mathbb{R}^{d \times d}$ and the gradient vector $\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta}))) \in \mathbb{R}^{d \times 1}$ exist almost surely $\forall k \in \{1, 2, \dots, n\}$ with respect to the probability measure \mathbb{P} ;
- (N3) $J_k(\mathbf{X}_k, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, $\forall k = 1, 2, \dots, n$, almost surely with respect to \mathbb{P} and is a measurable function of \mathbf{X}_k ;
- (N4) $\mathbb{E}[\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta}))) | \boldsymbol{\theta}] = \mathbf{0}$, $k = 1, 2, \dots, n$;
- (N5) with $[\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta})))]^\top \in \mathbb{R}^{1 \times d}$ denoting the transpose of $\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta})))$,

$$\mathbb{E}[[\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta})))][\nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta})))]^\top | \boldsymbol{\theta}] = -\mathbb{E}[J_k(\mathbf{X}_k, \boldsymbol{\theta})].$$

From now on we denote by $I_k(\boldsymbol{\theta}) := -\mathbb{E}[J_k(\mathbf{X}_k, \boldsymbol{\theta})]$;

(N6) for

$$\bar{I}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{j=1}^n I_j(\boldsymbol{\theta}), \quad (3)$$

there exists a matrix $\bar{I}(\boldsymbol{\theta}) \in \mathbb{R}^{d \times d}$ such that $\bar{I}_n(\boldsymbol{\theta}) \xrightarrow[n \rightarrow \infty]{} \bar{I}(\boldsymbol{\theta})$. In addition, $\bar{I}_n(\boldsymbol{\theta})$, $\bar{I}(\boldsymbol{\theta})$ are symmetric matrices for all $\boldsymbol{\theta}$ and $\bar{I}(\boldsymbol{\theta})$ is positive definite;

- (N7) for some $\delta > 0$, $\frac{\sum_k \mathbb{E}[\boldsymbol{\lambda}^\top \nabla(\log(f_k(\mathbf{X}_k | \boldsymbol{\theta}_0)))^{2+\delta}]}{n^{\frac{2+\delta}{2}}} \xrightarrow[n \rightarrow \infty]{} 0$ for all $\boldsymbol{\lambda} \in \mathbb{R}^d$;

- (N8) with $\|\cdot\|$ the ordinary Euclidean norm on \mathbb{R}^d , then for $k, i, j \in \{1, 2, \dots, d\}$ there exist $\epsilon > 0$, $K > 0$, $\delta > 0$ and random variables $B_{k,ij}(\mathbf{X}_k)$ such that

$$(i) \sup \left\{ \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f_k(\mathbf{X}_k | \mathbf{t})) \right| : \|\mathbf{t} - \boldsymbol{\theta}_0\| \leq \epsilon \right\} \leq B_{k,ij}(\mathbf{X}_k);$$

$$(ii) \ E |B_{k,ij}(\mathbf{X}_k)|^{1+\delta} \leq K.$$

Assuming that $\hat{\theta}_n(\mathbf{X})$ exists and is unique, the following theorem gives the result for the asymptotic normality of the MLE in the case of i.n.i.d. random vectors in a slightly different way than Hoadley (1971).

Theorem 2.1. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be independent random vectors with probability density (or mass) functions $f_i(\mathbf{x}_i|\theta)$, where $\theta \in \Theta \subset \mathbb{R}^d$. Assume that the MLE exists and is unique and that the regularity conditions (N1)-(N8) hold. Also let $\mathbf{Z} \sim N_d(\mathbf{0}, I_{d \times d})$, where $\mathbf{0}$ is the $d \times 1$ zero vector and $I_{d \times d}$ is the $d \times d$ identity matrix. Then, for $\bar{I}_n(\theta)$ as in (3)*

$$\sqrt{n} [\bar{I}_n(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{X}) - \theta_0) \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z}. \quad (4)$$

Proof. Hoadley (1971) proves in Theorem 2, p.1983 that under the regularity conditions (N1)-(N8)

$$\sqrt{n} (\hat{\theta}_n(\mathbf{X}) - \theta_0) \xrightarrow[n \rightarrow \infty]{d} [\bar{I}(\theta_0)]^{-\frac{1}{2}} \mathbf{Z}.$$

Using this result and (N6) we obtain that

$$[\bar{I}_n(\theta_0)]^{\frac{1}{2}} \sqrt{n} (\hat{\theta}_n(\mathbf{X}) - \theta_0) \xrightarrow[n \rightarrow \infty]{d} [\bar{I}(\theta_0)]^{\frac{1}{2}} [\bar{I}(\theta_0)]^{-\frac{1}{2}} \mathbf{Z} = \mathbf{Z},$$

which is the result of the theorem. ■

The normal approximation in (4) is an asymptotic result and since we will never have an infinite sample size, our motivation is to assess the quality of this normal approximation through explicit, for finite sample size, upper bounds on the distributional distance of interest. From now on, unless otherwise stated, $\bar{I}_n(\theta)$ is as in (3). Let the subscript (m) denote an index for which $|\hat{\theta}_n(\mathbf{x})_{(m)} - \theta_{0,(m)}|$ is the largest among the d components;

$$(m) \in \{1, 2, \dots, d\} \text{ is such that } |\hat{\theta}_n(\mathbf{x})_{(m)} - \theta_{0,(m)}| \geq |\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}|, \quad \forall j \in \{1, 2, \dots, d\}$$

and also, for ease of presentation, let

$$Q_{(m)} = Q_{(m)}(\mathbf{X}, \theta_0) := \hat{\theta}_n(\mathbf{X})_{(m)} - \theta_{0,(m)}. \quad (5)$$

Our main result is as follows.

Theorem 2.2. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be i.n.i.d. \mathbb{R}^t -valued, $t \in \mathbb{Z}^+$, random vectors with probability density (or mass) function $f_i(\mathbf{x}_i|\theta)$, for which the parameter space Θ is an open subset of \mathbb{R}^d . Assume that the MLE exists and is unique and that (N1)-(N8) are satisfied. In addition, assume that for any $\theta_0 \in \Theta$ there exists $0 < \epsilon = \epsilon(\theta_0)$ and functions $M_{kjl}(\mathbf{x})$, $\forall k, j, l \in \{1, 2, \dots, d\}$ such that $\left| \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_l} l(\theta, \mathbf{x}) \right| \leq M_{kjl}(\mathbf{x})$ for all $\theta \in \Theta$ such that $|\theta_j - \theta_{0,j}| < \epsilon$ $\forall j \in \{1, 2, \dots, d\}$. Also, for $Q_{(m)}$ as in (5), assume that $E \left((M_{kju}(\mathbf{X}))^2 | Q_{(m)}| < \epsilon \right) < \infty$. Let $\{\mathbf{X}'_i, i = 1, 2, \dots, n\}$ be an independent copy of $\{\mathbf{X}_i, i = 1, 2, \dots, n\}$. For $h \in H$, with H as in (1) and $\mathbf{Z} \sim N_d(\mathbf{0}, I_{d \times d})$ it holds that*

$$\begin{aligned} & \left| E \left[h \left(\sqrt{n} [\bar{I}_n(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{X}) - \theta_0) \right) \right] - E[h(\mathbf{Z})] \right| \\ & \leq \frac{\|h\|_1}{\sqrt{n}} K_1(\theta_0) + \frac{\|h\|_2}{\sqrt{n}} K_2(\theta_0) + \frac{\|h\|_3}{\sqrt{n}} K_3(\theta_0) + \frac{2\|h\|}{\epsilon^2} E \left(\sum_{j=1}^d (\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j})^2 \right), \end{aligned} \quad (6)$$

where

$$\begin{aligned}
K_1(\boldsymbol{\theta}_0) &= \sum_{k=1}^d \sum_{l=1}^d \left| [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \left| \sum_{j=1}^d \sqrt{\mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \mathbb{E} \left[\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{X}) + n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj} \right)^2 \right]} \right| \\
&\quad + \frac{1}{2} \left\{ \sum_{k=1}^d \sum_{l=1}^d \left| [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \left| \sum_{j=1}^d \sum_{v=1}^d \left[\mathbb{E} \left(\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \left(\hat{\theta}_n(\mathbf{X})_v - \theta_{0,v} \right)^2 \right) \right]^{\frac{1}{2}} \right. \right. \\
&\quad \left. \left. \times \left[\mathbb{E} \left((M_{kjv}(\mathbf{X}))^2 \mid |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\}, \tag{7}
\end{aligned}$$

$$\begin{aligned}
K_2(\boldsymbol{\theta}_0) &= \frac{1}{4\sqrt{n}} \sum_{j=1}^d \left[\sum_{i=1}^n \text{Var} \left(\left(\sum_{k=1}^d [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right)_{jk} \frac{\partial}{\partial \theta_k} \log(f_i(\mathbf{X}_i | \boldsymbol{\theta}_0)) \right)^2 \right]^{\frac{1}{2}} \\
&\quad + \frac{1}{2\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^d \left[\sum_{i=1}^n \text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right)_{jq} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{kv} \\
&\quad \left. \times \frac{\partial}{\partial \theta_q} \log(f_i(\mathbf{X}_i | \boldsymbol{\theta}_0)) \frac{\partial}{\partial \theta_v} \log(f_i(\mathbf{X}_i | \boldsymbol{\theta}_0)) \right)^{\frac{1}{2}} \tag{8}
\end{aligned}$$

and

$$K_3(\boldsymbol{\theta}_0) = \frac{1}{12n} \sum_{i=1}^n \mathbb{E} \left(\sum_{m=1}^d \left| \sum_{l=1}^d [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{ml} \left(\frac{\partial}{\partial \theta_l} \{ \log(f_i(\mathbf{X}'_i | \boldsymbol{\theta}_0)) - \log(f_i(\mathbf{X}_i | \boldsymbol{\theta}_0)) \} \right) \right)^3. \tag{9}$$

Outline of the proof. The regularity conditions and the definition of the MLE give that $\frac{\partial}{\partial \theta_k} l(\hat{\boldsymbol{\theta}}_n(\mathbf{x}); \mathbf{x}) = 0 \ \forall k \in \{1, 2, \dots, d\}$. A second-order Taylor expansion of $\frac{\partial}{\partial \theta_k} l(\hat{\boldsymbol{\theta}}_n(\mathbf{x}); \mathbf{x})$ about $\boldsymbol{\theta}_0$ yields

$$\begin{aligned}
&\sum_{j=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) \left(\frac{\partial^2}{\partial \theta_k \partial \theta_j} l(\boldsymbol{\theta}_0; \mathbf{x}) \right) \\
&= -\frac{\partial}{\partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{x}) - \frac{1}{2} \sum_{j=1}^d \sum_{q=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) (\hat{\theta}_n(\mathbf{x})_q - \theta_{0,q}) \left(\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_q} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right),
\end{aligned}$$

with $\boldsymbol{\theta}_0^*$ between $\hat{\boldsymbol{\theta}}_n(\mathbf{x})$ and $\boldsymbol{\theta}_0$. Adding $\sum_{j=1}^d n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj}(\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j})$ on both sides of the above equation gives that

$$\begin{aligned}
\sum_{j=1}^d n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj}(\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) &= \frac{\partial}{\partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{x}) + \sum_{j=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) \left(\frac{\partial^2}{\partial \theta_k \partial \theta_j} l(\boldsymbol{\theta}_0; \mathbf{x}) + n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj} \right) \\
&\quad + \frac{1}{2} \sum_{j=1}^d \sum_{q=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) (\hat{\theta}_n(\mathbf{x})_q - \theta_{0,q}) \left(\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_q} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right). \tag{10}
\end{aligned}$$

Using (10), which holds $\forall k \in \{1, 2, \dots, d\}$,

$$\begin{aligned} & \sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{x}) - \boldsymbol{\theta}_0) \\ &= \frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \left\{ \nabla(l(\boldsymbol{\theta}_0; \mathbf{x})) + \sum_{j=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) \left(\nabla \left(\frac{\partial}{\partial \theta_j} l(\boldsymbol{\theta}_0; \mathbf{x}) \right) + n [\bar{I}_n(\boldsymbol{\theta}_0)]_{[j]} \right) \right\} \\ &+ \frac{1}{2\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \left\{ \sum_{j=1}^d \sum_{q=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j}) (\hat{\theta}_n(\mathbf{x})_q - \theta_{0,q}) \left(\nabla \left(\frac{\partial^2}{\partial \theta_j \partial \theta_q} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right) \right) \right\}, \end{aligned}$$

where $[\bar{I}_n(\boldsymbol{\theta}_0)]_{[j]}$ is the j^{th} column of the matrix $\bar{I}_n(\boldsymbol{\theta}_0)$. The triangle inequality gives

$$\begin{aligned} & \left| \mathbb{E} \left[h \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| \\ & \leq \left| \mathbb{E} \left[h \left(\frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \nabla(l(\boldsymbol{\theta}_0; \mathbf{X})) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| \end{aligned} \quad (11)$$

$$+ \left| \mathbb{E} \left[h \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) - h \left(\frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \nabla(l(\boldsymbol{\theta}_0; \mathbf{X})) \right) \right] \right|. \quad (12)$$

Now, (11) is based on $\nabla(l(\boldsymbol{\theta}_0; \mathbf{x})) = \sum_{i=1}^n \nabla(\log(f_i(\mathbf{x}_i | \boldsymbol{\theta}_0)))$ which is a sum of independent random vectors. For this expression, a bound using Stein's method for multivariate normal approximation will be derived. In contrast, (12) will be bounded using multivariate Taylor expansions. Technical difficulties arise as the third-order partial derivatives of the log-likelihood function may not be uniformly bounded in $\boldsymbol{\theta}$. Therefore, for $0 < \epsilon = \epsilon(\boldsymbol{\theta}_0)$ we will condition on whether $|Q_{(m)}|$ as defined in (5) is greater or less than the positive constant ϵ and each case will be treated separately by bounding conditional expectations. Known probability inequalities, such as the Cauchy-Schwarz and Markov's inequality, will be employed in order to derive the upper bounds in each case.

Remark 2.1. (1) At first glance, the bound seems complicated. However, the examples that follow show that the terms are easily calculated giving an expression for the bound, which is of the optimal $n^{-\frac{1}{2}}$ -order.

(2) Assuming that $\bar{I}_n(\boldsymbol{\theta}_0) = \mathcal{O}(1)$ in (3) and using Theorem 2.1 yields, for fixed d , $\mathbb{E} \left(\sum_{j=1}^d (\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j})^2 \right) = \mathcal{O}(\frac{1}{n})$. To see this, use that from the asymptotic normality of the MLE as expressed in Theorem 2.1, $\sqrt{n} \mathbb{E}(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \xrightarrow{n \rightarrow \infty} \mathbf{0}$ and thus

$$\mathbb{E}(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j}) = o\left(\frac{1}{\sqrt{n}}\right), \quad \forall j \in \{1, 2, \dots, d\}. \quad (13)$$

Theorem 2.1 shows that $\text{Cov} \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) \xrightarrow{n \rightarrow \infty} I_{d \times d}$. Therefore,

$$n [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \text{Cov}(\hat{\boldsymbol{\theta}}_n(\mathbf{X})) [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \xrightarrow{n \rightarrow \infty} I_{d \times d}. \quad (14)$$

Assuming that the matrix $\bar{I}_n(\boldsymbol{\theta}_0)$ as defined in (3) is $\mathcal{O}(1)$, it follows from (14) that

$$\text{Var}(\hat{\theta}_n(\mathbf{X})_j) = \mathcal{O}\left(\frac{1}{n}\right), \quad \forall j \in \{1, 2, \dots, d\}.$$

Combining these results,

$$\mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] = \text{Var} \left(\hat{\theta}_n(\mathbf{X})_j \right) + \left[\mathbb{E} \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \right]^2 = \mathcal{O} \left(\frac{1}{n} \right). \quad (15)$$

Furthermore, using (13), (15) then if $\bar{I}_n(\boldsymbol{\theta}_0) = \mathcal{O}(1)$ it can be deduced that

$$K_1(\boldsymbol{\theta}_0) = \mathcal{O}(1), \quad K_2(\boldsymbol{\theta}_0) = \mathcal{O}(1), \quad K_3(\boldsymbol{\theta}_0) = \mathcal{O}(1), \quad (16)$$

where $K_1(\boldsymbol{\theta}_0), K_2(\boldsymbol{\theta}_0), K_3(\boldsymbol{\theta}_0)$ are as in (7), (8), (9), respectively. Hence, using (15) and (16), if $\bar{I}_n(\boldsymbol{\theta}_0) = \mathcal{O}(1)$ then the upper bound in Theorem 2.2 is $\mathcal{O} \left(\frac{1}{\sqrt{n}} \right)$.

(3) Often Hölder's inequality will be used to bound the third term as the calculation of absolute third moments can be quite complicated, even for simple multi-parameter distributions.

(4) In terms of the dimensionality d of the parameter, $K_1(\boldsymbol{\theta}_0) = \mathcal{O}(d^4)$, $K_2(\boldsymbol{\theta}_0) = \mathcal{O}(d^4)$ and $K_3(\boldsymbol{\theta}_0) = \mathcal{O}(d^8)$ as can be deduced from (7), (8) and (9), respectively. The last term of the bound in (6) is of order d in terms of the dimensionality of the parameter. Thus, for $d \gg n$ the bound does not behave well, but d could grow moderately with n . For example $d = o(n^\alpha)$, $0 < \alpha < \frac{1}{16}$ would still yield a bound which goes to zero as n goes to infinity.

2.2 Linear regression

This subsection calculates the bound in (6) for linear regression models. The asymptotic normality of the MLE in linear regression models has been proven in Fahrmeir and Kaufmann (1985). We give the example of a straight-line regression and the bound turns out to be, as expected, of order $\mathcal{O} \left(\frac{1}{\sqrt{n}} \right)$, where n is the sample size. The following notation is used throughout this subsection. The vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top \in \mathbb{R}^{n \times 1}$ denotes the response variable for the linear regression, while $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)^\top \in \mathbb{R}^{d \times 1}$ is the vector of the d parameters and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top \in \mathbb{R}^{n \times 1}$ is the vector of the error terms, which are i.i.d. random variables with $\epsilon_i \sim \mathcal{N}(0, \sigma^2) \forall i \in \{1, 2, \dots, n\}$. The true value of the unknown parameter $\boldsymbol{\beta}$ is denoted by $\boldsymbol{\beta}_0 = (\beta_{0,1}, \beta_{0,2}, \dots, \beta_{0,d})^\top \in \mathbb{R}^{d \times 1}$. The design matrix is

$$X = \begin{pmatrix} 1 & x_{1,2} & \dots & x_{1,d} \\ 1 & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,2} & \dots & x_{n,d} \end{pmatrix}.$$

For the model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

the aim is to find bounds on the distributional distance between the distribution of the MLE, $\hat{\boldsymbol{\beta}}$, and the normal distribution. The probability density function for Y_i is

$$f_i(y_i|\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - X_{[i]}\boldsymbol{\beta})^2 \right\}, \quad (17)$$

where $X_{[i]}$ denotes the i^{th} row of the design matrix. The parameter space $\Theta = \mathbb{R}^d$ is open and if $X^\top X$ is of full rank, the matrix $X^\top X$ is invertible and the vector MLE is

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{Y}. \quad (18)$$

We now bound the corresponding distributional distance.

Corollary 2.1. Let $Y_i, i \in \{1, 2, \dots, n\}$ be independent normal random variables with

$$Y_i \sim N(X_{[i]}\beta_0, \sigma^2),$$

where σ^2 is known. Assume that the $d \times d$ matrix $X^\top X$ is of full rank. Let $\{Y'_i, i = 1, 2, \dots, n\}$ be an independent copy of $\{Y_i, i = 1, 2, \dots, n\}$ and $\mathbf{Z} \sim N_d(\mathbf{0}, I_{d \times d})$ and $\bar{I}_n(\beta)$ is as in (3). Then for $h \in H$ as in (1)

$$\begin{aligned} & \left| E \left[h \left(\sqrt{n} [\bar{I}_n(\beta_0)]^{\frac{1}{2}} (\hat{\beta} - \beta_0) \right) \right] - E[h(\mathbf{Z})] \right| \\ & \leq \frac{\|h\|_2}{4} \sum_{j=1}^d \left[\sum_{i=1}^n \text{Var} \left(\left(\sum_{k=1}^d \frac{X_{ik}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right)_{jk} \left(Y_i - \sum_{m=1}^d X_{im} \beta_{0,m} \right) \right)^2 \right]^{\frac{1}{2}} \\ & + \frac{\|h\|_2}{2} \sum_{j>k}^d \sum_{k=1}^{d-1} \left[\sum_{i=1}^n \text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d \frac{X_{iq} X_{iv}}{\sigma^2} [X^\top X]^{-\frac{1}{2}} \right)_{jq} [X^\top X]^{-\frac{1}{2}}_{kv} \left(Y_i - \sum_{m=1}^d X_{im} \beta_{0,m} \right)^2 \right]^{\frac{1}{2}} \\ & + \frac{\|h\|_3}{12} \sum_{i=1}^n E \left(\sum_{m=1}^d \left| \sum_{l=1}^d \frac{X_{il}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right|_{ml} (Y_i - Y'_i) \right)^3. \end{aligned} \quad (19)$$

Proof. Using (17) we get that the Hessian matrix for the log-likelihood function,

$$\mathbf{H}(\beta; \mathbf{y}) = \left\{ \frac{\partial^2}{\partial \beta_i \partial \beta_j} l(\beta; \mathbf{y}) \right\}_{i,j=1,\dots,d} = -\frac{1}{\sigma^2} X^\top X,$$

does not depend on \mathbf{y} . Thus, $\bar{I}_n(\beta_0) = \frac{1}{n\sigma^2} X^\top X$ and so $[\bar{I}_n(\beta_0)]^{-\frac{1}{2}} = \sigma \sqrt{n} [X^\top X]^{-\frac{1}{2}}$. The result in (18) yields

$$\begin{aligned} \sqrt{n} [\bar{I}_n(\beta_0)]^{\frac{1}{2}} (\hat{\beta} - \beta_0) &= \frac{1}{\sigma} \left\{ [X^\top X]^{-\frac{1}{2}} X^\top \mathbf{Y} - [X^\top X]^{\frac{1}{2}} \beta_0 \right\} \\ &= \frac{1}{\sqrt{n}} \left[\sigma \sqrt{n} [X^\top X]^{-\frac{1}{2}} \right] \frac{1}{\sigma^2} (X^\top \mathbf{Y} - X^\top X \beta_0) \\ &= \frac{1}{\sqrt{n}} [I_n(\beta_0)]^{-\frac{1}{2}} \frac{d}{d\beta} l(\beta; \mathbf{y}) \Big|_{\beta=\beta_0}. \end{aligned} \quad (20)$$

The expression in (20) is the same as \mathbf{W} in (37) and therefore the quantity of interest $\left| E \left[h \left(\sqrt{n} [\bar{I}_n(\beta)]^{\frac{1}{2}} (\hat{\beta} - \beta_0) \right) \right] - E[h(\mathbf{Z})] \right|$ is equal to (11), with (12) being equal to zero for this specific case of the linear regression model. Thus, using (44) and

$$\frac{\partial}{\partial \beta_k} \log(f_i(Y_i | \beta_0)) = \frac{X_{ik}}{\sigma^2} \left(Y_i - \sum_{m=1}^d X_{im} \beta_{0,m} \right)$$

in Theorem 2.2 yields the result of the corollary. ■

Example: The simple linear model ($d=2$)

Here, we apply the results of (19) to the case of a straight-line regression with two unknown parameters. The model is

$$Y_i = \beta_1 + \beta_2(x_i - \bar{x}) + \epsilon_i, \quad \forall i \in \{1, 2, \dots, n\}.$$

The unknown parameters β_1 and β_2 are the *intercept* and *slope* of the regression, respectively. As before, the i.i.d. random variables $\epsilon_i \sim N(0, \sigma^2), \forall i \in \{1, 2, \dots, n\}$. The MLE exists, it is unique and equal to $\hat{\beta} = \left(\bar{Y}, \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^\top$.

Corollary 2.2. *Let Y_1, Y_2, \dots, Y_n be independent random variables with $Y_i \sim N(\beta_1 + \beta_2(x_i - \bar{x}), \sigma^2)$. The case of $x_i = x_j, \forall i, j \in \{1, 2, \dots, n\}$ with $i \neq j$ is excluded and for $\mathbf{Z} \sim N_2(\mathbf{0}, I_{2 \times 2})$ and $h \in H$ as in (1),*

$$\begin{aligned} \left| \mathbb{E} \left[h \left(\sqrt{n} [\bar{I}_n(\beta_0)]^{\frac{1}{2}} (\hat{\beta} - \beta_0) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| &\leq \frac{\|h\|_2}{4} \left(\sqrt{\frac{2}{n}} + \frac{\sqrt{2 \sum_{i=1}^n (x_i - \bar{x})^4}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) + \frac{\|h\|_2}{\sqrt{2n}} \\ &\quad + \frac{8\|h\|_3}{3\sqrt{\pi}} \left(\frac{1}{\sqrt{n}} + \frac{\sum_{i=1}^n |x_i - \bar{x}|^3}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{\frac{3}{2}}} \right). \end{aligned}$$

Proof. We have that

$$X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \quad X^\top X = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix}, \quad (X^\top X)^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \end{pmatrix}. \quad (21)$$

The result in (21) shows that $X^\top X$ is invertible if and only if $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$, which holds if x_i 's are not all identical. The quantities of the bound in (19) are calculated for this specific case. We use that $Y_i - \beta_1 - (x_i - \bar{x})\beta_2 \stackrel{d}{=} \sigma Z_i$, where $Z_i \sim N(0, 1)$. For the first term in (19) we obtain that

$$\begin{aligned} &\sum_{j=1}^2 \left[\sum_{i=1}^n \text{Var} \left(\left(\sum_{k=1}^2 \frac{X_{ik}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right)_{jk} \left(Y_i - \sum_{m=1}^2 X_{im} \beta_m \right) \right)^2 \right]^{\frac{1}{2}} \\ &= \sum_{j=1}^2 \left[\sum_{i=1}^n \text{Var} \left(\left(\left(\frac{X_{i1}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right)_{j1} + \frac{X_{i2}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right)_{j2} (\sigma Z_i) \right)^2 \right]^{\frac{1}{2}} \\ &= \frac{1}{n} \left[\sum_{i=1}^n \text{Var} (Z_i^2) \right]^{\frac{1}{2}} + \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \left[\sum_{i=1}^n (x_i - \bar{x})^4 \text{Var} (Z_i^2) \right]^{\frac{1}{2}} = \sqrt{\frac{2}{n}} + \frac{\sqrt{2 \sum_{i=1}^n (x_i - \bar{x})^4}}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned} \quad (22)$$

For an upper bound for the second term of (19), since $d = 2$ then $k = 1, j = 2$ leading to

$$\begin{aligned} &\left[\sum_{i=1}^n \text{Var} \left(\sum_{q=1}^2 \sum_{v=1}^2 \frac{X_{iq} X_{iv}}{\sigma^2} [X^\top X]^{-\frac{1}{2}} \right)_{2q} [X^\top X]^{-\frac{1}{2}} \left(Y_i - \sum_{m=1}^2 X_{im} \beta_m \right)^2 \right]^{\frac{1}{2}} \\ &= \left[\sum_{i=1}^n \text{Var} \left(\frac{X_{i2} X_{i1}}{\sigma^2} [X^\top X]^{-\frac{1}{2}} \right)_{22} [X^\top X]^{-\frac{1}{2}} (\sigma Z_i)^2 \right]^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}} \left[\sum_{i=1}^n \text{Var} (Z_i^2) (x_i - \bar{x})^2 \right]^{\frac{1}{2}} = \sqrt{\frac{2}{n}}. \end{aligned} \quad (23)$$

For the final term of (19), because Y'_i is an independent copy of Y_i , then $Y'_i - Y_i \sim N(0, 2\sigma^2)$, with $E|Y'_i - Y_i|^3 = 8\frac{\sigma^3}{\sqrt{\pi}}$. Using that

$$(|a| + |b|)^3 \leq 4(|a|^3 + |b|^3), \quad a, b \in \mathbb{R} \quad (24)$$

yields

$$\begin{aligned} & \sum_{i=1}^n E \left(\sum_{m=1}^2 \left| \sum_{l=1}^2 \frac{X_{il}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right|_{ml} (Y_i - Y'_i) \right)^3 \\ &= \sum_{i=1}^n E \left(\left| \left(\frac{X_{i1}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right)_{11} + \frac{X_{i2}}{\sigma} [X^\top X]^{-\frac{1}{2}} \right|_{22} (Y_i - Y'_i) \right)^3 \\ &\leq \sum_{i=1}^n E \left(\left(\frac{1}{\sigma\sqrt{n}} + \frac{|x_i - \bar{x}|}{\sigma\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \right) |Y'_i - Y_i| \right)^3 \\ &\leq 4 \sum_{i=1}^n \left(\frac{8}{n^{\frac{3}{2}}\sqrt{\pi}} + \frac{8|x_i - \bar{x}|^3}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{\frac{3}{2}}\sqrt{\pi}} \right) = \frac{32}{\sqrt{\pi}} \left(\frac{1}{\sqrt{n}} + \frac{\sum_{i=1}^n |x_i - \bar{x}|^3}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{\frac{3}{2}}} \right). \quad (25) \end{aligned}$$

Summarizing, in the case of Y_1, Y_2, \dots, Y_n being independent random variables with $Y_i \sim N(\beta_1 + \beta_2(x_i - \bar{x}), \sigma^2)$, we apply to (19) the results of (22), (23) and (25) to obtain the assertion of the corollary. \blacksquare

Remark 2.2. The bound is $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

2.3 Special case: Identically distributed random vectors

In this subsection we use weaker regularity conditions than (N1)-(N8) in order to find an upper bound in the case of independent and identically distributed random vectors. Following Davison (2008), we make the following assumptions:

- (R.C.1) The densities defined by any two different values of $\boldsymbol{\theta}$ are distinct;
- (R.C.2) the log-likelihood function is three times differentiable with respect to the unknown vector parameter, $\boldsymbol{\theta}$, and the third partial derivatives are continuous in $\boldsymbol{\theta}$.
- (R.C.3) for any $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and for \mathbb{X} denoting the support of the data $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, there exists $\epsilon_0 > 0$ and functions $M_{rst}(\mathbf{x})$ (they can depend on $\boldsymbol{\theta}_0$), such that for $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ and $r, s, t, j = 1, 2, \dots, d$,

$$\frac{1}{n} \left| \frac{\partial^3}{\partial \theta_r \partial \theta_s \partial \theta_t} l(\boldsymbol{\theta}; \mathbf{x}) \right| \leq M_{rst}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}, \quad |\theta_j - \theta_{0,j}| < \epsilon_0,$$

with $E(M_{rst}(\mathbf{X})) < \infty$;

- (R.C.4) the expected Fisher information matrix $I(\boldsymbol{\theta})$ is finite, symmetric and positive definite. For $r, s = 1, 2, \dots, d$, its elements satisfy

$$[I(\boldsymbol{\theta})]_{rs} = E \left\{ \frac{\partial}{\partial \theta_r} l(\boldsymbol{\theta}; \mathbf{X}) \frac{\partial}{\partial \theta_s} l(\boldsymbol{\theta}; \mathbf{X}) \right\} = E \left\{ -\frac{\partial^2}{\partial \theta_r \partial \theta_s} l(\boldsymbol{\theta}; \mathbf{X}) \right\}.$$

This condition implies that $I(\boldsymbol{\theta})$ is the covariance matrix of the score vector.

These regularity conditions in the multi-parameter case resemble those in [Anastasiou and Reinert \(2015\)](#) where it is assumed that the parameter is scalar. From now on, unless otherwise stated, the notation $I(\boldsymbol{\theta})$ stands for the expected Fisher information matrix for one random vector. Under (R.C.1)-(R.C.4), ([Davison, 2008](#), p.118) shows that

$$\sqrt{n}[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} N_d(\mathbf{0}, I_{d \times d}).$$

The upper bound on the distributional distance between the distribution of a vector MLE and the multivariate normal in the case of i.i.d. random vectors is the same as the bound in Theorem 2.2 and thus it is not given again. The bound can be simplified due to the fact that in the i.i.d. case $\bar{I}_n(\boldsymbol{\theta}_0) = I(\boldsymbol{\theta}_0)$ and $f_i(\mathbf{x}_i) = f(\mathbf{x}_i)$, $\forall i \in \{1, 2, \dots, n\}$. In the next example of independent random variables from the normal distribution with both mean and variance unknown the bound can be easily calculated and it is, as expected, of the order $\frac{1}{\sqrt{n}}$.

Example: The normal distribution

Here, we apply Theorem 2.2 in the case of X_1, X_2, \dots, X_n independent and identically distributed random variables from $N(\mu, \sigma^2)$ with $\boldsymbol{\theta}_0 = (\mu, \sigma^2)$. It is well-known that the MLE exists, it is unique and equal to $\hat{\boldsymbol{\theta}}_n(\mathbf{X}) = (\hat{\mu}, \hat{\sigma}^2)^\top = (\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2)^\top$; see for example [Davison \(2008\)](#), p.116. In addition, the regularity conditions (R.C.1)-(R.C.4) are satisfied. The proof of the following corollary is given in the Appendix.

Corollary 2.3. *Let X_1, X_2, \dots, X_n be i.i.d. random variables that follow the $N(\mu, \sigma^2)$ distribution. For $\mathbf{Z} \sim N_d(\mathbf{0}, I_{d \times d})$ and $h \in H$ as defined in (1),*

$$\begin{aligned} \left| \mathbb{E} \left[h \left(\sqrt{n}[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| &\leq \frac{5}{2} \frac{\|h\|_2}{\sqrt{n}} + 19 \frac{\|h\|_3}{\sqrt{n}} \\ &+ 8 \frac{\|h\|}{n\sigma^2} (1 + 2\sigma^2) + 4\sqrt{2} \frac{\|h\|_1}{\sqrt{n}} + 4 \frac{\|h\|_1}{\sqrt{n}} \left[\sqrt{2} + \sqrt{\frac{3}{2}} + 16\sqrt{2} \sqrt{\frac{1}{n} + \frac{\sigma^2}{4}} \right] \\ &+ 32 \frac{\|h\|_1}{\sqrt{n}} \left[1 + 648 \left[\left(\frac{3}{2} + \frac{\sigma^2}{4} \right)^2 + \frac{3}{n^2} \right] \right]^{\frac{1}{2}}. \end{aligned} \quad (26)$$

Remark 2.3. (1) The rate of convergence of the upper bound in (26) is $\frac{1}{\sqrt{n}}$.

(2) There might be cases where the parameters depend on the sample size, so that $\mu = \mu(n)$ and $\sigma^2 = \sigma^2(n)$. The bound in (26) does not depend on $\mu(n)$ and goes to zero as long as

$$(i) \quad \frac{1}{n\sigma^2(n)} \xrightarrow[n \rightarrow \infty]{} 0,$$

$$(ii) \quad \frac{\sigma^2(n)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} 0,$$

are both satisfied. From (i), the order of $\sigma^2(n)$ should not be less than or equal to $\frac{1}{n}$, while from (ii) we see that $\sigma^2(n)$ should be of order smaller than \sqrt{n} . For instance, $\sigma^2(n) = cn^{\frac{1}{4}}$, where $c \in \mathbb{R}$ is a constant, satisfies the above limits. The bound in (26) is then of order $\frac{1}{n^{\frac{1}{4}}}$.

3 Bounds when the MLE is not known explicitly

Anastasiou and Reinert (2015) give an upper bound for the mean squared error (MSE) of the MLE and use it to get upper bounds on the distributional distance of interest which can then be applied when the MLE is not expressed in a closed-form. In this section, we give similar bounds for the multi-parameter case with multivariate i.i.d. random vectors. We make some extra assumptions,

(Con.1) $\forall j \in \{1, 2, \dots, t\}$, the support S_j of X_{ij} is a bounded interval in \mathbb{R} ; let $s_j := \sup_{x_{ij} \in S_j} \{|x_{ij}|\}$ and $s := \max\{s_1, s_2, \dots, s_t\}$;

(Con.2) for all $\theta_0 \in \Theta$, where Θ is the open parameter space, there exists an $\epsilon_0 = \epsilon_0(\theta_0) > 0$ such that for all $\theta \in \Theta$ with $|\theta_j - \theta_{0,j}| < \epsilon_0, \forall j = 1, 2, \dots, d$

$$\sup_{\substack{\theta: |\theta_q - \theta_{0,q}| < \epsilon_0 \\ \forall q \in \{1, 2, \dots, d\}}} \left| \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_i} \log f(\mathbf{x}_1 | \theta) \right| \leq M_{kji},$$

where $M_{kji} = M_{kji}(\theta_0)$ is a constant that may depend only on θ_0 ;

(Con.3) for $M = \sup_{i,j} \left\{ \left| [I(\theta_0)]^{-1} \right|_{ij} \right\}$, the sample size satisfies

$$n > \frac{s^2 d^2}{4\epsilon_0^2} \left(M\epsilon_0 \sum_{l=1}^d \sum_{k=1}^d \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kim} \right. \\ \left. + \left[M^2 \epsilon_0^2 \left(\sum_{l=1}^d \sum_{k=1}^d \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kim} \right)^2 + 8M \right]^{\frac{1}{2}} \right)^2.$$

If (Con.1)-(Con.3) hold, then

$$2d^2 s^2 M + dsM \sqrt{n} \epsilon_0^2 \sum_{l=1}^d \sum_{k=1}^d \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kim} - n\epsilon_0^2 < 0$$

holds with ϵ_0 as in (Con.2). Section 2 gave an upper bound for the distributional distance between the distribution of the MLE and the multivariate normal distribution. As explained in the outline of the proof of Theorem 2.2, this bound in (6) can be split into terms coming from Stein's method, and terms due to Taylor expansions and conditional expectations. For ease of presentation, we abbreviate

$$D := D(\theta_0, h, \mathbf{X}) := \frac{\|h\|_2}{4\sqrt{n}} \sum_{j=1}^d \left[\text{Var} \left(\left(\sum_{k=1}^d [I(\theta_0)]^{-\frac{1}{2}} \right]_{jk} \frac{\partial}{\partial \theta_k} \log f(\mathbf{X}_1 | \theta_0) \right)^2 \right]^{\frac{1}{2}} \\ + \frac{\|h\|_2}{2\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^d \left[\text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d [I(\theta_0)]^{-\frac{1}{2}} \right]_{jq} \frac{\partial}{\partial \theta_q} \log f(\mathbf{X}_1 | \theta_0) [I(\theta_0)]^{-\frac{1}{2}} \right]_{kv} \frac{\partial}{\partial \theta_v} \log f(\mathbf{X}_1 | \theta_0) \right]^{\frac{1}{2}} \\ + \frac{\|h\|_3}{12\sqrt{n}} \mathbb{E} \left(\sum_{i=1}^d \left| \sum_{l=1}^d [I(\theta_0)]^{-\frac{1}{2}} \right|_{il} \left(\frac{\partial}{\partial \theta_l} \log f(\mathbf{X}'_1 | \theta_0) - \frac{\partial}{\partial \theta_l} \log f(\mathbf{X}_1 | \theta_0) \right) \right|^3. \quad (27)$$

In order to give an upper bound when $\hat{\theta}_n(\mathbf{X})$ is not known explicitly, we bound $\sqrt{\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]}$ by a quantity which does not require knowledge of the MLE. The result is given in Theorem 3.1 below, followed by a brief explanation of the idea of the proof. The complete proof is given in Section 4.

Theorem 3.1. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be i.i.d. \mathbb{R}^t -valued random elements, for $t \in \mathbb{N}$, with probability density (or mass) function $f(\mathbf{x}_i|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the d -valued vector parameter. Assume that (R.C.1)-(R.C.4) and also (Con.1)-(Con.3) are satisfied. Also $\mathbf{Z} \sim \mathcal{N}_d(\mathbf{0}, I_{d \times d})$ and we assume existence and uniqueness of the MLE, $\hat{\theta}_n(\mathbf{X})$. For $\epsilon = \epsilon_0$ as in (Con.2) and using the notation*

$$\begin{aligned} \gamma &= \sum_{j=1}^d \left| [I(\boldsymbol{\theta}_0)]^{-1} \right|_{jj} + \frac{M}{2\sqrt{n}} \sum_{j=1}^d \sqrt{\text{Var} \left(\left(\sum_{k=1}^d [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right)_{jk} \frac{\partial}{\partial \theta_k} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)^2} \\ &\quad + \frac{M}{\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^d \left[\text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right)_{jq} [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{kv} \frac{\partial}{\partial \theta_q} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_v} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right]^{\frac{1}{2}} \\ \omega &= 1 - 2 \frac{d^2 s^2 M}{n \epsilon^2} - \frac{dsM}{\sqrt{n}} \sum_{l=1}^d \sum_{k=1}^d \left| [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kmi} \\ v &= 2d^{\frac{3}{2}} sM \sum_{l=1}^d \sum_{k=1}^d \left| [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)} \end{aligned} \quad (28)$$

it holds that

$$\sqrt{\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]} \leq \frac{1}{\sqrt{n}} \left(\frac{\frac{v}{\sqrt{n}} + \sqrt{\frac{v^2}{n} + 4\omega\gamma}}{2\omega} \right) := U_1. \quad (29)$$

Idea of the proof. The general bound in Theorem 2.2 as expressed in (6) is based on terms the calculation of which requires an analytic expression of the MLE. Our first step is to use (Con.1)-(Con.3) in order to put the dependence of the bound on the MLE only through the MSE, $\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$. The aim now becomes to upper bound $\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$ by a quantity that does not contain any terms related to $\hat{\theta}_n(\mathbf{X})$. To this aim, we will first use the result of Theorem (2.2) for the test function

$$h = h_{\theta_0} : \mathbb{R}^d \rightarrow \mathbb{R} : h(\mathbf{x}) = \mathbf{x}^\top [I(\boldsymbol{\theta}_0)]^{-1} \mathbf{x}. \quad (30)$$

Notice that with B_h denoting the bound in Theorem (2.2) for h as in (30),

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] &= \frac{1}{n} \left| \mathbb{E} \left[h \left(\sqrt{n} [I(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0 \right) \right) \right] - \mathbb{E}[h(\mathbf{Z})] + \mathbb{E}[h(\mathbf{Z})] \right| \\ &\leq \frac{1}{n} (B_h + |\mathbb{E}[h(\mathbf{Z})]|). \end{aligned} \quad (31)$$

Since B_h is also an expression of $\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$ and its square root, then the final step of the proof is to solve a simple quadratic inequality with unknown the quantity $\sqrt{\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]}$, which will lead to the required bound for the MSE. This bound will not contain any terms related to $\hat{\boldsymbol{\theta}}_n(\mathbf{X})$.

Remark 3.1. (1) The quantities γ , ω and v are $\mathcal{O}(1)$ and therefore $U_1 = \mathcal{O} \left(\frac{1}{\sqrt{n}} \right)$, with U_1 as in (29).

(2) As the bound (29) does not include $\hat{\boldsymbol{\theta}}_n(\mathbf{X})$, in cases where a closed-form expression for the vector MLE is not available, we can still get an upper bound on the distributional distance between the distribution of the MLE and the d -variate standard normal. For D as in (27) and U_1 as in (29),

$$\begin{aligned} \left| \mathbb{E} \left[h \left(\sqrt{n} [I(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0 \right) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| &\leq D + 2 \frac{\|h\|}{\epsilon^2} U_1^2 \\ &+ \|h\|_1 \sqrt{d} U_1 \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \right)} \\ &+ \frac{\|h\|_1 \sqrt{n}}{2} U_1^2 \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^d \sum_{i=1}^d M_{kmi}. \end{aligned} \quad (32)$$

Example: The Beta distribution

Here, we find an upper bound for the specific example of i.i.d. random variables from the Beta distribution with both shape parameters being unknown. An analytic expression for the MLE is not available. Applying the result of (29) to bound $\mathbb{E} \left(\sum_{j=1}^2 \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right)$, gives an upper bound for the distributional distance of interest. Some useful notations are now presented. Firstly, $\Psi_j(\cdot)$ is the j^{th} derivative of the digamma function Ψ , with $\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$, $z > 0$. The function $\Psi_j(z)$ can be defined through a sum, with

$$\Psi_m(z) = (-1)^{m+1} m! \sum_{k=0}^{\infty} \frac{1}{(z+k)^{m+1}}, \text{ for } z \in \mathbb{C} \setminus \{\mathbb{Z}_0^-\} \text{ and } m > 0. \quad (33)$$

For $\alpha, \beta, x, y > 0$ and $0 < \epsilon < \min\{x, y\}$, let

$$\begin{aligned}
\delta_I &:= \Psi_1(\alpha)\Psi_1(\beta) - \Psi_1(\alpha + \beta)(\Psi_1(\alpha) + \Psi_1(\beta)), \\
C_1(x, y) &:= \Psi_3(x) + \Psi_3(x + y) + 3[\Psi_1(x)]^2 + 3[\Psi_1(x + y)]^2, \\
C_2(x, y) &:= \Psi_1(x) - \Psi_1(x + y) + \sqrt{\delta_I}, \\
C_3(x, y) &:= C_1(x, y)[C_2(y, x)]^2, \\
C_4(x, y) &:= \frac{6x}{(y - \epsilon)^4} + \frac{x\pi^4}{15} + \frac{6}{(x + y - \epsilon)^3} + 7.26, \\
M_B &:= \frac{1}{\delta_I} \sup\{\Psi_1(\alpha + \beta), \Psi_1(\min\{\alpha, \beta\}) - \Psi_1(\alpha + \beta)\}, \\
\gamma_B &:= \frac{4M_B}{\sqrt{n}\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ \left[(C_2(\beta, \alpha))^4 C_1(\alpha, \beta) + [\Psi_1(\alpha + \beta)]^4 C_1(\beta, \alpha) \right]^{\frac{1}{2}} \right. \\
&\quad \left. + \left[(C_2(\alpha, \beta))^4 C_1(\beta, \alpha) + [\Psi_1(\alpha + \beta)]^4 C_1(\alpha, \beta) \right]^{\frac{1}{2}} \right\} \\
&\quad + \frac{M_B \sqrt{24}}{\sqrt{n}\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ [\Psi_1(\alpha + \beta)]^2 (C_3(\alpha, \beta) + C_3(\beta, \alpha)) \right. \\
&\quad \left. + 2\sqrt{C_1(\alpha, \beta)C_1(\beta, \alpha)} [[\Psi_1(\alpha + \beta)]^4 + [C_2(\alpha, \beta)]^2 [C_2(\beta, \alpha)]^2] \right\}^{\frac{1}{2}} \\
&\quad + \frac{\Psi_1(\beta) + \Psi_1(\alpha) - 2\Psi_1(\alpha + \beta)}{\delta_I} \\
\omega_B &:= 1 - 8 \frac{M_B}{n\epsilon^2} - \frac{2M_B \{ (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta) \}}{\sqrt{n}\delta_I (C_2(\alpha, \beta) + C_2(\beta, \alpha))}. \tag{34}
\end{aligned}$$

Corollary 3.1 below gives the upper bound related to the Beta distribution. The proof is given in the Appendix.

Corollary 3.1. *Let X_1, X_2, \dots, X_n be i.i.d. random variables from the $\text{Beta}(\alpha, \beta)$ distribution with $\theta_0 = (\alpha, \beta)$. Let $m = \min\{\alpha, \beta\}$, $\epsilon = \frac{m}{2} > 0$ and*

$$\begin{aligned}
n &\geq \frac{4}{m^2} \left(\frac{mM_B (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta)}{2\sqrt{\delta_I} (C_2(\alpha, \beta) + C_2(\beta, \alpha))} \right. \\
&\quad \left. + \left[\left(\frac{mM_B (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta)}{2\sqrt{\delta_I} (C_2(\alpha, \beta) + C_2(\beta, \alpha))} \right)^2 + 8M_B \right]^{\frac{1}{2}} \right)^2. \tag{35}
\end{aligned}$$

Then,

$$\mathbf{a)} \text{ When } n \text{ satisfies (35), } \mathbb{E} \left[\sum_{j=1}^2 \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \leq \sqrt{\frac{\gamma_B}{n\omega_B}}.$$

b) For $\mathbf{Z} \sim N_2(\mathbf{0}, I_{2 \times 2})$, we obtain

$$\begin{aligned}
& \left| \mathbb{E} \left[h \left(\sqrt{n} [I(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0 \right) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| \leq \\
& \frac{2\|h\|_2}{\sqrt{n}\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ \left[(C_2(\beta, \alpha))^4 C_1(\alpha, \beta) + [\Psi_1(\alpha + \beta)]^4 C_1(\beta, \alpha) \right]^{\frac{1}{2}} \right. \\
& \quad \left. + \left[(C_2(\alpha, \beta))^4 C_1(\beta, \alpha) + [\Psi_1(\alpha + \beta)]^4 C_1(\alpha, \beta) \right]^{\frac{1}{2}} \right\} \\
& + \frac{\|h\|_2 \sqrt{6}}{\sqrt{n}\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ [\Psi_1(\alpha + \beta)]^2 (C_3(\alpha, \beta) + C_3(\beta, \alpha)) \right. \\
& \quad \left. + 2\sqrt{C_1(\alpha, \beta)C_1(\beta, \alpha)} \left[[\Psi_1(\alpha + \beta)]^4 + [C_2(\alpha, \beta)]^2 [C_2(\beta, \alpha)]^2 \right] \right\}^{\frac{1}{2}} \\
& + \frac{32\|h\|_3 8^{\frac{3}{4}}}{3\sqrt{n} [\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]]^{\frac{3}{2}}} \left\{ \left[[C_2(\beta, \alpha)]^3 + [\Psi_1(\alpha + \beta)]^3 \right] [C_1(\alpha, \beta)]^{\frac{3}{4}} \right. \\
& \quad \left. + \left[[C_2(\alpha, \beta)]^3 + [\Psi_1(\alpha + \beta)]^3 \right] [C_1(\beta, \alpha)]^{\frac{3}{4}} \right\} \\
& + \frac{8\|h\|\gamma_B}{nm^2\omega_B} + \frac{\|h\|_1\gamma_B \left\{ (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta) \right\}}{2\sqrt{n}\omega_B \sqrt{\delta_I (C_2(\alpha, \beta) + C_2(\beta, \alpha))}}. \tag{36}
\end{aligned}$$

Remark 3.2. Using the notation (34), it is straightforward that the first three terms of the bound are $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$. In addition, since γ_B and ω_B are $\mathcal{O}(1)$, the fourth and the fifth term of the bound are of order $\frac{1}{n}$ and $\frac{1}{\sqrt{n}}$, respectively. Combining these results for the order of each of the terms, the order of the bound (36) is $\frac{1}{\sqrt{n}}$.

4 Proofs of Theorems 2.2 and 3.1

In this section the complete steps of the proofs of the two main theorems of our paper are given. The following lemma (special case of Chebyshev's 'other' inequality) is useful for bounding conditional expectations, which sometimes can be difficult to derive. The proof is given in the Appendix.

Lemma 4.1. *Let $\mathbf{M} \in \mathbb{R}^d$ be a random vector with $M_i > 0 \ \forall i = 1, 2, \dots, d$ and $\epsilon > 0$. For every continuous function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $f(\mathbf{m})$ is increasing and $f(\mathbf{m}) \geq 0$, for $m_i > 0 \ \forall i \in \{1, 2, \dots, d\}$, where $\mathbf{m} = (m_1, m_2, \dots, m_d)$,*

$$\mathbb{E}[f(\mathbf{M}) | M_i < \epsilon \ \forall i = 1, 2, \dots, d] \leq \mathbb{E}[f(\mathbf{M})].$$

Proof of Theorem 2.2. It has already been shown in the outline of the proof in p.6 that the triangle inequality yields

$$\left| \mathbb{E} \left[h \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0 \right) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| \leq (11) + (12).$$

Step 1: Upper bound for (11). First of all, $\nabla(l(\boldsymbol{\theta}_0; \mathbf{x})) = \sum_{i=1}^n \nabla(\log(f_i(\mathbf{x}_i|\boldsymbol{\theta}_0)))$ due to independence. The results of Theorem 2.1 of Reinert and Röllin (2009) will be used for

$$\mathbf{W} = \frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \sum_{i=1}^n \nabla(\log(f_i(\mathbf{X}_i|\boldsymbol{\theta}_0))) = (W_1, W_2, \dots, W_d)^\top \in \mathbb{R}^{d \times 1}. \quad (37)$$

From (37) we have that for all $k \in \{1, 2, \dots, d\}$, $W_k = \sum_{i=1}^n \xi_{ik}$, with

$$\xi_{ik} = \frac{1}{\sqrt{n}} \sum_{j=1}^d \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{kj} \frac{\partial}{\partial \theta_j} \log(f_i(\mathbf{X}_i|\boldsymbol{\theta}_0)). \quad (38)$$

From the regularity conditions, $E(\nabla(l(\boldsymbol{\theta}_0; \mathbf{X}))) = \mathbf{0}$ and thus $E(\mathbf{W}) = \mathbf{0}$. Also, $\bar{I}_n(\boldsymbol{\theta}_0)$ is symmetric. Therefore, $[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}$ is also symmetric. Using the regularity conditions we know that $\sum_{i=1}^n \text{Cov}[\nabla(\log(f_i(\mathbf{X}_i|\boldsymbol{\theta}_0)))] = n\bar{I}_n(\boldsymbol{\theta}_0)$ and basic calculations show that $\text{Cov}(\mathbf{W}) = I_{d \times d}$. Since $E(\mathbf{W}) = \mathbf{0}$ and $E(\mathbf{W}\mathbf{W}^\top) = I_{d \times d}$, the first assumption of Theorem 2.1 from Reinert and Röllin (2009) is satisfied. This theorem also assumes that $\exists \mathbf{W}'$ such that $(\mathbf{W}, \mathbf{W}')$ consists an exchangeable pair meaning that $(\mathbf{W}, \mathbf{W}') \stackrel{d}{=} (\mathbf{W}', \mathbf{W})$, where $\stackrel{d}{=}$ denotes equality in distribution. In addition, it is assumed that

$$E(\mathbf{W}' - \mathbf{W} | \mathbf{W}) = -\Lambda \mathbf{W} + \mathbf{R} \quad (39)$$

for an invertible $d \times d$ matrix Λ and a $\sigma(\mathbf{W})$ -measurable random vector \mathbf{R} . To define \mathbf{W}' in our case such that (39) is satisfied, let $\{\mathbf{X}'_i, i = 1, 2, \dots, n\}$ be an independent copy of $\{\mathbf{X}_i, i = 1, 2, \dots, n\}$ and let the index $I \in \{1, 2, \dots, n\}$ follow the uniform distribution on $\{1, 2, \dots, n\}$, independently of $\{\mathbf{X}_i, \mathbf{X}'_i, i = 1, 2, \dots, n\}$. Let

$$\xi'_{ik} = \frac{1}{\sqrt{n}} \sum_{j=1}^d \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{kj} \frac{\partial}{\partial \theta_j} \log(f_i(\mathbf{X}'_i|\boldsymbol{\theta}_0))$$

and

$$W'_k = W_k - \xi_{Ik} + \xi'_{Ik}, \quad \forall k \in \{1, 2, \dots, d\},$$

with $E(W'_k - W_k | \mathbf{W}) = E(\xi'_{Ik} - \xi_{Ik} | \mathbf{W}) = -E(\xi_{Ik} | \mathbf{W}) = -\frac{1}{n} \sum_{i=1}^n E(\xi_{ik} | \mathbf{W}) = -\frac{W_k}{n}$. Hence (39) is satisfied with $\Lambda = \frac{1}{n} I_{d \times d}$ and $\mathbf{R} = \mathbf{0}$. Thus, Theorem 2.1 from Reinert and Röllin (2009) gives in our case that

$$|E[h(\mathbf{W})] - E[h(\mathbf{Z})]| \leq n \left(\frac{\|h\|_2}{4} \sum_{i=1}^d \sum_{j=1}^d [\text{Var}(E[(W'_i - W_i)(W'_j - W_j) | \mathbf{W}])]^{\frac{1}{2}} \right) \quad (40)$$

$$+ n \left(\frac{\|h\|_3}{12} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d E| (W'_i - W_i)(W'_j - W_j)(W'_k - W_k) | \right). \quad (41)$$

To bound the variance of the conditional expectations in (40), let $\mathcal{A} = \sigma(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$. Since $\sigma(\mathbf{W}) \subset \mathcal{A}$, for any random variable Y , $\text{Var}(E[Y | \mathbf{W}]) \leq \text{Var}(E[Y | \mathcal{A}])$. Then,

$$(40) \leq n \frac{\|h\|_2}{4} \left\{ \sum_{j=1}^d \sqrt{\text{Var}(E[(\xi'_{Ij} - \xi_{Ij})^2 | \mathcal{A}])} + 2 \sum_{j>k}^d \sum_{k=1}^{d-1} \sqrt{\text{Var}(E[(\xi'_{Ik} - \xi_{Ik})(\xi'_{Ij} - \xi_{Ij}) | \mathcal{A}])} \right\}. \quad (42)$$

Since $\{\mathbf{X}'_i, i = 1, 2, \dots, n\}$ is an independent copy of $\{\mathbf{X}_i, i = 1, 2, \dots, n\}$ and ξ'_{ik} is independent of \mathcal{A} ,

$$(42) = n \frac{\|h\|_2}{4} \left\{ \sum_{j=1}^d [\text{Var}(\mathbb{E}[(\xi'_{Ij})^2] - 2\mathbb{E}[\xi'_{Ij}]\mathbb{E}[\xi_{Ij}|\mathcal{A}] + \mathbb{E}[\xi_{Ij}^2|\mathcal{A}])]^{\frac{1}{2}} + 2 \sum_{k=1}^{d-1} \sum_{j>k}^d [\text{Var}(\mathbb{E}[\xi'_{Ik}\xi'_{Ij}] - \mathbb{E}[\xi'_{Ij}]\mathbb{E}[\xi_{Ik}|\mathcal{A}] - \mathbb{E}[\xi'_{Ik}]\mathbb{E}[\xi_{Ij}|\mathcal{A}] + \mathbb{E}[\xi_{Ik}\xi_{Ij}|\mathcal{A}])]^{\frac{1}{2}} \right\}. \quad (43)$$

Using that $\mathbb{E}[\xi'_{ik}] = 0$,

$$(43) = n \frac{\|h\|_2}{4} \left\{ \sum_{j=1}^d \left[\frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n \mathbb{E}[\xi_{ij}^2|\mathcal{A}] \right) \right]^{\frac{1}{2}} + 2 \sum_{j>k}^d \sum_{k=1}^{d-1} \left[\frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n \mathbb{E}[\xi_{ik}\xi_{ij}|\mathcal{A}] \right) \right]^{\frac{1}{2}} \right\} \\ = \frac{\|h\|_2}{4} \left\{ \sum_{j=1}^d \left[\text{Var} \left(\sum_{i=1}^n \xi_{ij}^2 \right) \right]^{\frac{1}{2}} + 2 \sum_{j>k}^d \sum_{k=1}^{d-1} \left[\text{Var} \left(\sum_{i=1}^n \xi_{ik}\xi_{ij} \right) \right]^{\frac{1}{2}} \right\} = \frac{\|h\|_2}{\sqrt{n}} K_2(\boldsymbol{\theta}_0),$$

with $K_2(\boldsymbol{\theta}_0)$ defined in (8). For (41), using (38), after basic calculations we obtain that

$$(41) \leq \frac{\|h\|_3}{\sqrt{n}} K_3(\boldsymbol{\theta}_0),$$

with $K_3(\boldsymbol{\theta}_0)$ as in (9). Thus,

$$(11) \leq \frac{\|h\|_2}{\sqrt{n}} K_2(\boldsymbol{\theta}_0) + \frac{\|h\|_3}{\sqrt{n}} K_3(\boldsymbol{\theta}_0). \quad (44)$$

Step 2: Upper bound for (12). For ease of presentation, let

$$\mathbf{R}_1(\boldsymbol{\theta}_0; \mathbf{x}) = \frac{1}{2\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \sum_{j=1}^d \sum_{q=1}^d (\hat{\theta}_n(\mathbf{x})_j - \theta_{0,j})(\hat{\theta}_n(\mathbf{x})_q - \theta_{0,q}) \left(\nabla \left(\frac{\partial^2}{\partial \theta_j \partial \theta_q} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right) \right) \\ T_1 = T_1(\boldsymbol{\theta}_0; \mathbf{X}, h) := h \left(\sqrt{n} [\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) \\ - h \left(\frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} (\nabla(l(\boldsymbol{\theta}_0; \mathbf{x}))) + \mathbf{R}_1(\boldsymbol{\theta}_0; \mathbf{X}) \right) \\ T_2 = T_2(\boldsymbol{\theta}_0; \mathbf{X}, h) := h \left(\frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} (\nabla(l(\boldsymbol{\theta}_0; \mathbf{x}))) + \mathbf{R}_1(\boldsymbol{\theta}_0; \mathbf{x}) \right) \\ - h \left(\frac{1}{\sqrt{n}} [\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} (\nabla(l(\boldsymbol{\theta}_0; \mathbf{X}))) \right). \quad (45)$$

Using the above notation and the triangle inequality

$$(12) = |\mathbb{E}[T_1 + T_2]| \leq \mathbb{E}|T_1| + \mathbb{E}|T_2|.$$

With $A_{[j]}$ the j^{th} row of a matrix A , a first order multivariate Taylor expansion gives that

$$|T_1| \leq \|h\|_1 \left| \sum_{j=1}^d \left(\sqrt{n} \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \right]_{[j]} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) - \frac{1}{\sqrt{n}} \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{[j]} \nabla(l(\boldsymbol{\theta}_0; \mathbf{X})) \right. \right. \\ \left. \left. - \frac{1}{2\sqrt{n}} \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{[j]} \left\{ \sum_{k=1}^d \sum_{q=1}^d (\hat{\theta}_n(\mathbf{x})_k - \theta_{0,k}) (\hat{\theta}_n(\mathbf{x})_q - \theta_{0,q}) \right. \right. \right. \\ \left. \left. \left. \times \left(\nabla \left(\frac{\partial^2}{\partial \theta_k \partial \theta_q} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right) \right) \right\} \right) \right|.$$

Using (10) component-wise, we conclude that

$$\begin{aligned} \mathbb{E}[T_1] &\leq \frac{\|h\|_1}{\sqrt{n}} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \left| \sum_{j=1}^d \mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{X}) + n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj} \right) \right] \right| \right| \\ &\leq \frac{\|h\|_1}{\sqrt{n}} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \left| \sum_{j=1}^d \left[\mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \right. \right. \right. \\ &\quad \left. \left. \times \mathbb{E} \left[\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{X}) + n[\bar{I}_n(\boldsymbol{\theta}_0)]_{kj} \right)^2 \right] \right] \right|^{\frac{1}{2}} \right|, \end{aligned} \quad (46)$$

using Cauchy-Schwarz inequality. To bound now $\mathbb{E}[T_2]$, with T_2 as in (45), we take into account that $\frac{\partial^3}{\partial \theta_k \partial \theta_q \partial \theta_j} l(\boldsymbol{\theta}; \mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}$ is in general not uniformly bounded. For $\epsilon > 0$, the law of total expectation and Markov's inequality yield

$$\begin{aligned} \mathbb{E}[T_2] &\leq 2\|h\| \mathbb{P} \left(\left| \hat{\theta}_n(\mathbf{X})_{(m)} - \theta_{0,(m)} \right| \geq \epsilon \right) + \mathbb{E} \left(|T_2| \left| \left| \hat{\theta}_n(\mathbf{X})_{(m)} - \theta_{0,(m)} \right| < \epsilon \right) \\ &\leq \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left(\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right) + \mathbb{E} \left(|T_2| \left| \left| \hat{\theta}_n(\mathbf{X})_{(m)} - \theta_{0,(m)} \right| < \epsilon \right), \end{aligned} \quad (47)$$

To bound $\mathbb{E} \left(|T_2| \left| \left| \hat{\theta}_n(\mathbf{X})_{(m)} - \theta_{0,(m)} \right| < \epsilon \right)$, a first-order Taylor expansion and (10) yield

$$\begin{aligned} |T_2| &\leq \frac{\|h\|_1}{2\sqrt{n}} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \\ &\quad \times \left\{ \sum_{j=1}^d \sum_{v=1}^d \left| \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\hat{\theta}_n(\mathbf{X})_v - \theta_{0,v} \right) \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_v} l(\boldsymbol{\theta}; \mathbf{X}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right| \right\}. \end{aligned} \quad (48)$$

Therefore, from (47) and (48) we have, for $Q_{(m)}$ as in (5), that

$$\begin{aligned} \mathbb{E}|T_2| &\leq \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left(\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right) \\ &+ \frac{\|h\|_1}{2\sqrt{n}} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \mathbb{E} \left(\sum_{j=1}^d \sum_{v=1}^d \left| \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\hat{\theta}_n(\mathbf{X})_v - \theta_{0,v} \right) \right. \right. \\ &\quad \left. \left. \times \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_v} l(\boldsymbol{\theta}; \mathbf{X}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right| \Big| Q_{(m)} \Big| < \epsilon \right). \end{aligned}$$

The Cauchy-Schwarz inequality and Lemma 4.1 yield

$$\begin{aligned} \mathbb{E}|T_2| &\leq \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left(\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right) \\ &+ \frac{\|h\|_1}{2\sqrt{n}} \left\{ \sum_{k=1}^d \sum_{l=1}^d \left| \left[[\bar{I}_n(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{j=1}^d \sum_{v=1}^d \left[\mathbb{E} \left(\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \left(\hat{\theta}_n(\mathbf{X})_v - \theta_{0,v} \right)^2 \right) \right]^{\frac{1}{2}} \right. \\ &\quad \left. \times \left[\mathbb{E} \left((M_{k j v}(\mathbf{X}))^2 \Big| Q_{(m)} \right) < \epsilon \right]^{\frac{1}{2}} \right\}. \end{aligned} \quad (49)$$

Therefore, from (46) and (49) we obtain that

$$(12) \leq \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left(\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right) + \frac{\|h\|_1}{\sqrt{n}} K_1(\boldsymbol{\theta}_0), \quad (50)$$

where $K_1(\boldsymbol{\theta}_0)$ is as in (7). Using now (44) and (50) we obtain the assertion. \blacksquare

Proof of Theorem 3.1. Using (6) and (27) and with $Q_{(m)}$ as in (5),

$$\begin{aligned} \left| \mathbb{E} \left[h \left(\sqrt{n} [I(\boldsymbol{\theta}_0)]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_n(\mathbf{X}) - \boldsymbol{\theta}_0) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| &\leq D + \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \\ &+ \frac{\|h\|_1}{\sqrt{n}} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{j=1}^d \left[\mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \mathbb{E} \left[\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{X}) + n[I(\boldsymbol{\theta}_0)]_{kj} \right)^2 \right] \right]^{\frac{1}{2}} \end{aligned} \quad (51)$$

$$\begin{aligned} &+ \frac{\|h\|_1}{2\sqrt{n}} \left\{ \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \right. \\ &\quad \left. \times \mathbb{E} \left(\sum_{j=1}^d \sum_{i=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right) \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_i} l(\boldsymbol{\theta}_0^*; \mathbf{X}) \right) \Big| \Big| Q_{(m)} \Big| < \epsilon \right\}. \end{aligned} \quad (52)$$

Step 1: Upper bound for (51). Since $E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}l(\boldsymbol{\theta}_0; \mathbf{X}) + n[I(\boldsymbol{\theta}_0)]_{kj}\right) = 0, \forall j, k \in \{1, 2, \dots, d\}$,

$$\begin{aligned}
(51) &= \|h\|_1 \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{j=1}^d \sqrt{E \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]} \sqrt{\text{Var} \left(\frac{\partial^2}{\partial\theta_k\partial\theta_j} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)} \\
&\leq \|h\|_1 \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{j=1}^d \sqrt{E \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]} \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial\theta_k\partial\theta_i} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)},
\end{aligned} \tag{53}$$

where the inequality comes from the trivial bound

$$\text{Var} \left(\frac{\partial^2}{\partial\theta_k\partial\theta_j} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right) \leq \sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial\theta_k\partial\theta_i} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)$$

since the variance of a random variable is always non-negative. Now, using that $\left(\sum_j^d \alpha_j\right)^2 \leq d \left(\sum_{j=1}^d \alpha_j^2\right)$ for $\alpha_j \in \mathbb{R}$, yields

$$\left(\sum_{j=1}^d \sqrt{E \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]} \right)^2 \leq d \sum_{j=1}^d E \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right].$$

Taking square roots in both sides of the above inequality and applying this result to (53) gives

$$\begin{aligned}
(51) &\leq \|h\|_1 \sqrt{d} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{lk} \right| \\
&\quad \times \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial\theta_k\partial\theta_i} \log f(\mathbf{X}_1|\boldsymbol{\theta}_0) \right)} \sqrt{E \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]}.
\end{aligned} \tag{54}$$

Step 2: Upper bound for (52). Notice that from (Con.2), if $\left| \hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right| < \epsilon$, $\forall j \in \{1, 2, \dots, d\}$, then $\left| \frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i} l(\boldsymbol{\theta}_0^*; \mathbf{x}) \right| = \left| \sum_{l=1}^n \frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i} \log f(\mathbf{x}_l|\boldsymbol{\theta}_0^*) \right| \leq n M_{kji}$. Also,

$$\begin{aligned}
&\sum_{j=1}^d \sum_{i=1}^d \left| \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right) \right| M_{kji} \\
&= \sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 M_{kjj} + 2 \sum_{j>i}^d \sum_{i=1}^{d-1} \left| \hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right| \left| \hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right| M_{kij}.
\end{aligned}$$

Using now that $2\alpha\beta \leq \alpha^2 + \beta^2, \forall \alpha, \beta \in \mathbb{R}$,

$$\begin{aligned}
&\sum_{j=1}^d \sum_{i=1}^d \left| \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right) \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right) \right| M_{kji} \\
&\leq \sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 M_{kjj} + \sum_{j>i}^d \sum_{i=1}^{d-1} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 + \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right)^2 \right] M_{kji} \\
&= \sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \sum_{i=1}^d M_{kji} \leq \sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \sum_{m=1}^d \sum_{i=1}^d M_{kmi}.
\end{aligned} \tag{55}$$

Using (55) yields

$$(52) \leq \frac{\|h\|_1 \sqrt{n}}{2} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^d \sum_{i=1}^d M_{kmi} \mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]. \quad (56)$$

Hence, from (54) and (56),

$$\begin{aligned} & \left| \mathbb{E} \left[h \left(\sqrt{n} [I(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{X}) - \theta_0) \right) \right] - \mathbb{E}[h(\mathbf{Z})] \right| \leq D + \frac{2\|h\|}{\epsilon^2} \mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \\ & + \|h\|_1 \sqrt{d} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\mathbf{X}_1 | \theta_0) \right)} \sqrt{\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]} \\ & + \frac{\|h\|_1 \sqrt{n}}{2} \sum_{k=1}^d \sum_{l=1}^d \left| \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^d \sum_{i=1}^d M_{kmi} \mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]. \end{aligned} \quad (57)$$

Since D as defined in (27), is not related to the MLE, the upper bound in (57) depends on $\hat{\theta}_n(\mathbf{X})$ only through $\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$. Our purpose is to find a bound for $\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$ that does not contain any terms related to $\hat{\theta}_n(\mathbf{X})$.

Step 3: The MSE test function. To this purpose define the test function h as in (30). Then, since $\left([I(\theta_0)]^{\frac{1}{2}} \right)^\top = [I(\theta_0)]^{\frac{1}{2}}$ as $[I(\theta_0)]^{\frac{1}{2}}$ is symmetric, we have that

$$\begin{aligned} h \left(\sqrt{n} [I(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{x}) - \theta_0) \right) &= \left[\sqrt{n} [I(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{x}) - \theta_0) \right]^\top [I(\theta_0)]^{-1} \sqrt{n} [I(\theta_0)]^{\frac{1}{2}} (\hat{\theta}_n(\mathbf{x}) - \theta_0) \\ &= n \left(\hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^\top \left(\hat{\theta}_n(\mathbf{X}) - \theta_0 \right) = n \sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2. \end{aligned}$$

Hence,

$$\mathbb{E}[h(\mathbf{Z})] = \sum_{j=1}^d \left[[I(\theta_0)]^{-1} \right]_{jj} \mathbb{E}[Z_j^2] + 2 \sum_{\substack{k=1 \\ k < i}}^d \sum_{i=1}^d \left[[I(\theta_0)]^{-1} \right]_{ki} \mathbb{E}[Z_k] \mathbb{E}[Z_i] = \sum_{j=1}^d \left[[I(\theta_0)]^{-1} \right]_{jj}.$$

Denoting by B_h the bound in (57) for the test function h as in (30), we get using (31) that

$$\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \leq \frac{1}{n} \left(B_h + \sum_{j=1}^d \left| \left[[I(\theta_0)]^{-1} \right]_{jj} \right| \right). \quad (58)$$

For the calculation of B_h , note that with s as in (Con.1) and M as in (Con.3),

$$\begin{aligned} \|h\| &= d^2 s^2 M, & \|h\|_1 &= \sup_j \left| \frac{\partial}{\partial x_j} h(\mathbf{x}) \right| = 2dMs, \\ \|h\|_2 &= \sup_{i,j} \left| \frac{\partial^2}{\partial x_i \partial x_j} h(\mathbf{x}) \right| = 2M, & \|h\|_3 &= \sup_{i,j,k} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h(\mathbf{x}) \right| = 0. \end{aligned} \quad (59)$$

For $U := \sqrt{\mathbb{E} \left[\sum_{j=1}^d \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]}$, the results in (27), (57) and (59) yield

$$\begin{aligned}
B_h &= \frac{M}{2\sqrt{n}} \sum_{j=1}^d \left[\text{Var} \left(\left(\sum_{k=1}^d \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right)_{jk} \frac{\partial}{\partial \theta_k} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \right)^2 \right]^{\frac{1}{2}} \\
&+ \frac{M}{\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^d \left[\text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right)_{jq} \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right)_{kv} \frac{\partial}{\partial \theta_q} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_v} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \right]^{\frac{1}{2}} \\
&+ U^2 \left(2 \frac{d^2 s^2 M}{\epsilon^2} + \sqrt{n} ds M \sum_{l=1}^d \sum_{k=1}^d \left| \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kmi} \right) \\
&+ 2U d^{\frac{3}{2}} s M \sum_{l=1}^d \sum_{k=1}^d \left| \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right|_{lk} \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \right)}. \tag{60}
\end{aligned}$$

The results in (58) and (60) give that

$$\begin{aligned}
0 &\leq U^2 \left(2 \frac{d^2 s^2 M}{n \epsilon^2} + \frac{dsM}{\sqrt{n}} \sum_{l=1}^d \sum_{k=1}^d \left| \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right|_{lk} \sum_{m=1}^d \sum_{i=1}^d M_{kmi} - 1 \right) \\
&+ U \left(\frac{2}{n} d^{\frac{3}{2}} s M \sum_{l=1}^d \sum_{k=1}^d \left| \left[I(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right|_{lk} \sqrt{\sum_{i=1}^d \text{Var} \left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\mathbf{X}_1 | \boldsymbol{\theta}_0) \right)} \right) + \frac{\gamma}{n}, \tag{61}
\end{aligned}$$

with γ as in (28). Solving the quadratic inequality in (61) (with unknown U) and using (Con.3) related to the sample size, n , we obtain for v and ω as in (28) that

$$U \leq U_1 = \frac{1}{\sqrt{n}} \left(\frac{\frac{v}{\sqrt{n}} + \sqrt{\frac{v^2}{n} + 4\omega\gamma}}{2\omega} \right), \tag{62}$$

proving the result of the theorem. ■

Acknowledgements

The author would like to thank Gesine Reinert for various insightful comments and suggestions. Andreas Anastasiou was supported by a Teaching Assistantship Bursary from the Department of Statistics, University of Oxford, and the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K503113/1.

Appendix: Proofs of Lemma 4.1 and of Corollaries 2.3 and 3.1

Proof of Lemma 4.1. Let $k \in \{1, 2, \dots, d\}$. We set $M_{d+1} = 0$. It will be shown that for $k = 1, 2, \dots, d$ we have that

$$\mathbb{E}[f(\mathbf{M}) | M_i < \epsilon, i = k, \dots, d] \leq \mathbb{E}[f(\mathbf{M}) | M_i < \epsilon, i = k+1, \dots, d].$$

From the law of total expectation,

$$\begin{aligned} & \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d] \\ &= \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k, \dots, d] \mathbb{P}[M_k < \epsilon | M_i < \epsilon, i = k+1, \dots, d] \\ &+ \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d, M_k \geq \epsilon] \mathbb{P}[M_k \geq \epsilon | M_i < \epsilon, i = k+1, \dots, d]. \end{aligned}$$

Using that

$$\mathbb{P}[M_k < \epsilon | M_i < \epsilon, i = k+1, \dots, d] = 1 - \mathbb{P}[M_k \geq \epsilon | M_i < \epsilon, i = k+1, \dots, d]$$

yields

$$\begin{aligned} & \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d] \\ &= \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k, \dots, d] \\ &+ \mathbb{P}[M_k \geq \epsilon | M_i < \epsilon, i = k+1, \dots, d] \left\{ \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d, M_k \geq \epsilon] \right. \\ &\quad \left. - \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k, \dots, d] \right\}. \end{aligned} \quad (63)$$

Since $f(\mathbf{m})$ is an increasing function,

$$\mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d, M_k \geq \epsilon] - \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k, \dots, d] \geq 0. \quad (64)$$

Applying this to (63) gives that

$$\mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k, \dots, d] \leq \mathbb{E}[f(\mathbf{M})|M_i < \epsilon, i = k+1, \dots, d].$$

A simple iteration over k gives that

$$\mathbb{E}[f(\mathbf{M})|M_i < \epsilon \forall i = 1, 2, \dots, d] \leq \mathbb{E}[f(\mathbf{M})], \quad (65)$$

which is the result of the lemma. ■

Proof of Corollary 2.3. The first and second-order partial derivatives are

$$\begin{aligned} \frac{\partial}{\partial \mu} l(\boldsymbol{\theta}_0; \mathbf{x}) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), & \frac{\partial}{\partial \sigma^2} l(\boldsymbol{\theta}_0; \mathbf{x}) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2, \\ \frac{\partial^2}{\partial \mu^2} l(\boldsymbol{\theta}_0; \mathbf{x}) &= -\frac{n}{\sigma^2}, & \frac{\partial^2}{\partial \sigma^4} l(\boldsymbol{\theta}_0; \mathbf{x}) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2, \\ \frac{\partial^2}{\partial \mu \partial \sigma^2} l(\boldsymbol{\theta}_0; \mathbf{x}) &= \frac{\partial^2}{\partial \sigma^2 \partial \mu} l(\boldsymbol{\theta}_0; \mathbf{x}) = -\frac{n}{\sigma^4} (\bar{X} - \mu). \end{aligned} \quad (66)$$

Hence, the expected Fisher Information matrix for one random variable is

$$I(\boldsymbol{\theta}_0) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \text{ so that } [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} = \begin{pmatrix} \sigma & 0 \\ 0 & \sqrt{2}\sigma^2 \end{pmatrix}. \quad (67)$$

Using (37), let $\mathbf{W} = \frac{1}{\sqrt{n}} \left(\frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu), -\frac{n}{\sqrt{2}} + \frac{1}{\sqrt{2}\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right)^\top = (W_1, W_2)^\top$, where $W_i = \sum_{j=1}^n \xi_{ji}$ with $\xi_{i1} = \frac{X_i - \mu}{\sqrt{n}\sigma}$ and $\xi_{i2} = \frac{(X_i - \mu)^2 - \sigma^2}{\sigma^2 \sqrt{2n}}$, $i = 1, 2, \dots, n$. We bound the terms in Theorem 2.2 in order of appearance. The term $K_1(\boldsymbol{\theta}_0)$ is given in (7) and for the first quantity of $\frac{\|h\|_1}{\sqrt{n}} K_1(\boldsymbol{\theta}_0)$, using that

$$\mathbb{E} \left[(\bar{X} - \mu)^2 \right] = \frac{\sigma^2}{n}, \quad \mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^2 \right] = \frac{\sigma^4}{n} \left(2 - \frac{1}{n} \right) \quad (68)$$

yields

$$\begin{aligned} & \frac{\|h\|_1}{\sqrt{n}} \sum_{k=1}^2 \sum_{l=1}^2 \left| \left[\bar{I}_n(\boldsymbol{\theta}_0) \right]^{-\frac{1}{2}} \right|_{lk} \left| \sum_{j=1}^2 \left[\mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right] \mathbb{E} \left[\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}_0; \mathbf{X}) + n [\bar{I}_n(\boldsymbol{\theta}_0)]_{kj} \right)^2 \right] \right] \right|^{\frac{1}{2}} \\ &= \frac{\sigma \|h\|_1}{\sqrt{n}} \left[\mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^2 \right] \mathbb{E} \left[\left(\frac{-n(\bar{X} - \mu)}{\sigma^4} \right)^2 \right] \right]^{\frac{1}{2}} \\ &+ \frac{\sqrt{2}\sigma^2 \|h\|_1}{\sqrt{n}} \left\{ \left[\mathbb{E} \left[(\bar{X} - \mu)^2 \right] \mathbb{E} \left[\left(\frac{-n(\bar{X} - \mu)}{\sigma^4} \right)^2 \right] \right]^{\frac{1}{2}} \right. \\ &\quad \left. + \left[\mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^2 \right] \mathbb{E} \left[\left(\frac{n}{\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (X_i - \mu)^2 \right)^2 \right] \right]^{\frac{1}{2}} \right\} \\ &\leq \frac{\sqrt{n} \|h\|_1}{\sigma^3} \sqrt{\mathbb{E} \left(\hat{\sigma}^2 - \sigma^2 \right)^2} \sqrt{\mathbb{E} \left(\bar{X} - \mu \right)^2} + \frac{\sqrt{2n} \|h\|_1}{\sigma^2} \mathbb{E} \left[(\bar{X} - \mu)^2 \right] \\ &+ \frac{\sqrt{2}\sigma^2 \|h\|_1}{\sqrt{n}} \sqrt{\mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^2 \right] \mathbb{E} \left[\left(\frac{n}{\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (X_i - \mu)^2 \right)^2 \right]} \\ &= \frac{\|h\|_1}{\sqrt{n}} \sqrt{2 - \frac{1}{n}} + \frac{\sqrt{2} \|h\|_1}{\sqrt{n}} + \frac{\sqrt{2}\sigma^4 \|h\|_1}{n} \sqrt{\left(2 - \frac{1}{n} \right) \left(\frac{n^2}{\sigma^8} - \frac{2n^2}{\sigma^8} + \frac{n}{\sigma^8} (2 + n) \right)} \\ &\leq 4\sqrt{2} \frac{\|h\|_1}{\sqrt{n}} \quad (69) \end{aligned}$$

For the second quantity in (7), we find an upper bound for

$$\sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_i} l(\boldsymbol{\theta}; \mathbf{X}) \right| \leq n M_{kji}(\mathbf{X}),$$

where $m, k, j, i \in \{1, 2\}$ and $M_{kji}(\mathbf{x})$ is as in (R.C.3). Below, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is the vector parameter and $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}) = (\mu, \sigma^2)$ is the true, unknown value. We have

$$\sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_1^3} l(\boldsymbol{\theta}; \mathbf{X}) \right| = 0 =: M_{111}(\mathbf{X})$$

as well as

$$\begin{aligned}
\sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_2^3} l(\boldsymbol{\theta}; \mathbf{X}) \right| &= \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| -\frac{n}{\theta_2^3} + \frac{3}{\theta_2^4} \sum_{i=1}^n (X_i - \theta_1)^2 \right| \\
&< \frac{n}{(\sigma^2 - \epsilon)^3} + \frac{3}{(\sigma^2 - \epsilon)^4} \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu + \mu - \theta_1)^2 \right| \\
&\leq \frac{n}{(\sigma^2 - \epsilon)^3} + \frac{3}{(\sigma^2 - \epsilon)^4} \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| 3 \sum_{i=1}^n \left[(X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + (\mu - \theta_1)^2 \right] \right| \\
&< \frac{n}{(\sigma^2 - \epsilon)^3} + \frac{9n}{(\sigma^2 - \epsilon)^4} \left(\hat{\sigma}^2 + (\bar{X} - \mu)^2 + \epsilon^2 \right) =: nM_{222}(\mathbf{X}).
\end{aligned} \tag{70}$$

Moreover,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_1 \partial \theta_2^2} l(\boldsymbol{\theta}; \mathbf{X}) \right| &= \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_2^2 \partial \theta_1} l(\boldsymbol{\theta}; \mathbf{X}) \right| = \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{2n}{\theta_2^3} (\bar{X} - \theta_1) \right| \\
&= \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{2n}{\theta_2^3} (\bar{X} - \mu + \mu - \theta_1) \right| < \frac{2n}{(\sigma^2 - \epsilon)^3} (|\bar{X} - \mu| + \epsilon) =: nM_{122}(\mathbf{X})
\end{aligned} \tag{71}$$

and

$$\begin{aligned}
\sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_1^2 \partial \theta_2} l(\boldsymbol{\theta}; \mathbf{X}) \right| &= \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{\partial^3}{\partial \theta_2 \partial \theta_1^2} l(\boldsymbol{\theta}; \mathbf{X}) \right| \\
&= \sup_{\boldsymbol{\theta}: |\theta_m - \theta_{0,m}| < \epsilon} \left| \frac{n}{\theta_2^2} \right| < \frac{n}{(\sigma^2 - \epsilon)^2} =: nM_{112}(\mathbf{X}).
\end{aligned} \tag{72}$$

In addition, we calculate $E \left(\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right)^2 \right), \forall i, j \in \{1, 2\}$. For this purpose we use that in the case of i.i.d. random variables from the Normal distribution,

$$\text{Cov} \left(\bar{X} - \mu, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right) = 0 \quad (\text{Casella and Berger, 2002, p.218}). \tag{73}$$

We have that $E(\bar{X} - \mu)^4 = \frac{3\sigma^4}{n^2}$ and $E \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)^4 = \frac{\sigma^8}{n^2} \left(12 + \frac{4}{n} - \frac{15}{n^2} \right) < 16 \frac{\sigma^8}{n^2}$, for all $n \in \mathbb{N}$. With $G \sim \chi_{n-1}^2$, so that $E(G) = n - 1$ and $\text{Var}(G) = 2(n - 1)$,

$$\begin{aligned}
&E \left[(\bar{X} - \mu)^2 \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)^2 \right] \\
&= E \left[(\bar{X} - \mu)^2 \right] E \left[\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)^2 \right] \quad (\text{ using (73)}) \\
&= E \left[(\bar{X} - \mu)^2 \right] \frac{\sigma^4}{n^2} E \left[\left(\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 - n \right)^2 \right] \\
&= \frac{\sigma^6}{n^3} E \left[(G - (n - 1) - 1)^2 \right] = \frac{\sigma^6}{n^3} (E \left[(G - (n - 1))^2 \right] - 2E[G - (n - 1)] + 1) \\
&= \frac{\sigma^6}{n^3} (\text{Var}[G] + 1) = \frac{\sigma^6}{n^3} (2(n - 1) + 1) = \frac{\sigma^6}{n^2} \left(2 - \frac{1}{n} \right) < 2 \frac{\sigma^6}{n^2}.
\end{aligned}$$

For $Q_{(m)}$ as in (5) and $[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}$ in (67), the second quantity in $K_1(\boldsymbol{\theta}_0)$ becomes

$$\begin{aligned}
& \frac{\|h\|_1}{2\sqrt{n}} \left\{ \sum_{k=1}^2 \sum_{l=1}^2 \left| [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \sum_{j=1}^2 \sum_{i=1}^2 \left[\mathbb{E} \left(\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right)^2 \right) \right]^{\frac{1}{2}} \right. \\
& \quad \left. \times \left[\mathbb{E} \left((nM_{kji}(\mathbf{X}))^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\} \\
& = \frac{\|h\|_1 \sigma}{2\sqrt{n}} \left\{ \frac{2n}{(\sigma^2 - \epsilon)^2} \left[\mathbb{E} \left((\bar{X} - \mu)^2 (\hat{\sigma}^2 - \sigma^2)^2 \right) \right]^{\frac{1}{2}} \right. \\
& \quad \left. + \left[\mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^4 \right] \right]^{\frac{1}{2}} \frac{2n}{(\sigma^2 - \epsilon)^3} \left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\} \\
& + \frac{\|h\|_1 \sqrt{2} \sigma^2}{2\sqrt{n}} \left\{ \frac{n}{(\sigma^2 - \epsilon)^2} \left[\mathbb{E} (\bar{X} - \mu)^4 \right]^{\frac{1}{2}} \right. \\
& \quad + \frac{4n}{(\sigma^2 - \epsilon)^3} \left[\mathbb{E} \left((\bar{X} - \mu)^2 (\hat{\sigma}^2 - \sigma^2)^2 \right) \right]^{\frac{1}{2}} \left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \\
& \quad \left. + n \left[\mathbb{E} \left[(\hat{\sigma}^2 - \sigma^2)^4 \right] \right]^{\frac{1}{2}} \left[\mathbb{E} \left((M_{222}(\mathbf{X}))^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\} \\
& \leq \frac{\|h\|_1 \sigma}{2\sqrt{n}} \left\{ \frac{2\sqrt{2} \sigma^3}{(\sigma^2 - \epsilon)^2} + \frac{8\sigma^4}{(\sigma^2 - \epsilon)^3} \left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\} \\
& + \frac{\|h\|_1 \sigma^2}{\sqrt{2n}} \left\{ \frac{\sqrt{3} \sigma^2}{(\sigma^2 - \epsilon)^2} + 4\sqrt{2} \frac{\sigma^3}{(\sigma^2 - \epsilon)^3} \left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right. \\
& \quad \left. + 4\sigma^4 \left[\mathbb{E} \left((M_{222}(\mathbf{X}))^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\}. \tag{74}
\end{aligned}$$

We next bound $\left[\mathbb{E} \left((M_{222}(\mathbf{X}))^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}}$ and $\left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}}$. For $\left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}}$, we have

$$\begin{aligned}
\left[\mathbb{E} \left((|\bar{X} - \mu| + \epsilon)^2 \middle| |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} & \leq \left[2\mathbb{E} \left((\bar{X} - \mu)^2 \middle| |Q_{(m)}| < \epsilon \right) + 2\epsilon^2 \right]^{\frac{1}{2}} \\
& \leq \sqrt{2} \left[\mathbb{E} \left[(\bar{X} - \mu)^2 \right] + \epsilon^2 \right]^{\frac{1}{2}} = \sqrt{2} \sqrt{\frac{\sigma^2}{n} + \epsilon^2}, \tag{75}
\end{aligned}$$

where for the second inequality we employed Lemma 4.1. In addition, simple steps yield

$$\begin{aligned}
M_{222}(\mathbf{X}) & = \frac{1}{(\sigma^2 - \epsilon)^3} + \frac{9}{(\sigma^2 - \epsilon)^4} \left(\hat{\sigma}^2 + (\bar{X} - \mu)^2 + \epsilon^2 \right) \\
& \leq \frac{1}{(\sigma^2 - \epsilon)^3} + \frac{9}{(\sigma^2 - \epsilon)^4} \left(|\hat{\sigma}^2 - \sigma^2| + (\bar{X} - \mu)^2 + \epsilon^2 + \sigma^2 \right),
\end{aligned}$$

which leads to

$$\begin{aligned} [M_{222}(\mathbf{X})]^2 &\leq 2 \left[\frac{1}{(\sigma^2 - \epsilon)^6} + \frac{81}{(\sigma^2 - \epsilon)^8} \left(|\hat{\sigma}^2 - \sigma^2| + (\bar{X} - \mu)^2 + \epsilon^2 + \sigma^2 \right)^2 \right] \\ &\leq 2 \left[\frac{1}{(\sigma^2 - \epsilon)^6} + \frac{162}{(\sigma^2 - \epsilon)^8} \left[\left(|\hat{\sigma}^2 - \sigma^2| + \epsilon^2 + \sigma^2 \right)^2 + (\bar{X} - \mu)^4 \right] \right]. \end{aligned} \quad (76)$$

Using the result in (76) and Lemma 4.1 yields

$$\left[\mathbb{E} \left((M_{222}(\mathbf{X}))^2 \mid |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \leq \sqrt{2} \left[\frac{1}{(\sigma^2 - \epsilon)^6} + \frac{162}{(\sigma^2 - \epsilon)^8} \left[(\epsilon + \epsilon^2 + \sigma^2)^2 + 3 \frac{\sigma^4}{n^2} \right] \right]^{\frac{1}{2}}. \quad (77)$$

Using (74), (75) and (77), the second term in (7) multiplied by $\frac{\|h\|_1}{\sqrt{n}}$ becomes

$$\begin{aligned} &\frac{\|h\|_1}{2\sqrt{n}} \left\{ \sum_{k=1}^d \sum_{l=1}^d \left| [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right|_{lk} \left| \sum_{j=1}^d \sum_{i=1}^d \left[\mathbb{E} \left(\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \left(\hat{\theta}_n(\mathbf{X})_i - \theta_{0,i} \right)^2 \right) \right]^{\frac{1}{2}} \right. \right. \\ &\quad \left. \left. \times \left[\mathbb{E} \left((nM_{kji}(\mathbf{X}))^2 \mid |Q_{(m)}| < \epsilon \right) \right]^{\frac{1}{2}} \right\} \\ &\leq \frac{\|h\|_1 \sigma^4}{\sqrt{n} (\sigma^2 - \epsilon)^2} \left[\sqrt{2} + \frac{4\sqrt{2}\sigma}{\sigma^2 - \epsilon} \sqrt{\frac{\sigma^2}{n} + \epsilon^2} \right] + \frac{\|h\|_1 \sigma^4}{\sqrt{2n} (\sigma^2 - \epsilon)^2} \left[\sqrt{3} + \frac{8\sigma}{\sigma^2 - \epsilon} \sqrt{\frac{\sigma^2}{n} + \epsilon^2} \right] \\ &\quad + \frac{4\|h\|_1 \sigma^6}{\sqrt{n} (\sigma^2 - \epsilon)^3} \left[1 + \frac{162}{(\sigma^2 - \epsilon)^2} \left[(\epsilon + \epsilon^2 + \sigma^2)^2 + 3 \frac{\sigma^4}{n^2} \right] \right]^{\frac{1}{2}} \\ &= \frac{\|h\|_1 \sigma^4}{\sqrt{n} (\sigma^2 - \epsilon)^2} \left[\sqrt{2} + \sqrt{\frac{3}{2}} + \frac{8\sqrt{2}\sigma}{\sigma^2 - \epsilon} \sqrt{\frac{\sigma^2}{n} + \epsilon^2} \right] \\ &\quad + \frac{4\|h\|_1 \sigma^6}{\sqrt{n} (\sigma^2 - \epsilon)^3} \left[1 + \frac{162}{(\sigma^2 - \epsilon)^2} \left[(\epsilon + \epsilon^2 + \sigma^2)^2 + 3 \frac{\sigma^4}{n^2} \right] \right]^{\frac{1}{2}}. \end{aligned} \quad (78)$$

Taking $\epsilon = \frac{\sigma^2}{2}$ yields

$$(78) = \frac{4\|h\|_1}{\sqrt{n}} \left[\sqrt{2} + \sqrt{\frac{3}{2}} + 16\sqrt{2} \sqrt{\frac{1}{n} + \frac{\sigma^2}{4}} \right] + \frac{32\|h\|_1}{\sqrt{n}} \left[1 + 648 \left[\left(\frac{3}{2} + \frac{\sigma^2}{4} \right)^2 + \frac{3}{n^2} \right] \right]^{\frac{1}{2}}. \quad (79)$$

Combining the results of (69) and (79) yields to a bound on the first term, $\frac{\|h\|_1}{\sqrt{n}} K_1(\boldsymbol{\theta}_0)$, of the general upper bound in (6). The second term, $\frac{\|h\|_2}{\sqrt{n}} K_2(\boldsymbol{\theta}_0)$, is a sum of two quantities as (8)

shows. For the first quantity, since $\frac{X_1 - \mu}{\sigma} \stackrel{d}{=} \sigma Z$, where $Z \sim N(0, 1)$,

$$\begin{aligned} &\frac{\|h\|_2}{4\sqrt{n}} \sum_{j=1}^d \left[\text{Var} \left(\left(\sum_{k=1}^d [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right)_{jk} \frac{\partial}{\partial \theta_k} \log f(X_1 | \boldsymbol{\theta}_0) \right)^2 \right]^{\frac{1}{2}} \\ &= \frac{\|h\|_2}{4\sqrt{n}} \left\{ \left[\text{Var} \left(\frac{(X_1 - \mu)^2}{\sigma^2} \right) \right]^{\frac{1}{2}} + \left[\text{Var} \left(\frac{(X_1 - \mu)^4}{2\sigma^4} - \frac{(X_1 - \mu)^2}{\sigma^2} \right) \right]^{\frac{1}{2}} \right\} \\ &= \frac{\|h\|_2}{4\sqrt{n}} \left\{ \sqrt{2} + \left[\frac{1}{4} \text{Var}(Z^4) + \text{Var}(Z^2) - \text{Cov}(Z^4, Z^2) \right]^{\frac{1}{2}} \right\} = \frac{\|h\|_2}{4\sqrt{n}} (\sqrt{2} + \sqrt{14}). \end{aligned} \quad (80)$$

For the second quantity in $\frac{\|h\|_2}{\sqrt{n}} K_2(\theta_0)$, since $[I(\theta_0)]^{-\frac{1}{2}}$ is diagonal, we obtain that

$$\begin{aligned}
& \frac{\|h\|_2}{2\sqrt{n}} \sum_{j>k}^d \sum_{k=1}^{d-1} \left[\text{Var} \left(\sum_{q=1}^d \sum_{v=1}^d \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{jq} \frac{\partial}{\partial \theta_q} \log f(X_1 | \theta_0) \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{kv} \frac{\partial}{\partial \theta_v} \log f(X_1 | \theta_0) \right) \right]^{\frac{1}{2}} \\
&= \frac{\|h\|_2}{2\sqrt{n}} \left[\text{Var} \left(\left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{11} \frac{\partial}{\partial \mu} \log f(X_1 | \theta_0) \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{22} \frac{\partial}{\partial \sigma^2} \log f(X_1 | \theta_0) \right) \right]^{\frac{1}{2}} \\
&= \frac{\|h\|_2}{2\sqrt{n}} \left[\text{Var} \left(\left(\frac{X_1 - \mu}{\sigma} \right) \left(\frac{(X_1 - \mu)^2}{\sqrt{2}\sigma^2} - \frac{1}{\sqrt{2}} \right) \right) \right]^{\frac{1}{2}} \\
&= \frac{\|h\|_2}{2\sqrt{2n}} [\text{Var}(Z^3 - Z)]^{\frac{1}{2}} = \frac{\|h\|_2}{2\sqrt{2n}} [\text{Var}(Z^3) + \text{Var}(Z) - 2\text{Cov}(Z^3, Z)]^{\frac{1}{2}} = \frac{\sqrt{5}\|h\|_2}{2\sqrt{n}}. \quad (81)
\end{aligned}$$

For an upper bound for the third term $\frac{\|h\|_3}{\sqrt{n}} K_3(\theta_0)$ in (6), with $K_3(\theta_0)$ as in (9), we use that X'_1 is an independent copy of X_1 . The triangle inequality and (24) give that

$$\begin{aligned}
& \frac{\|h\|_3}{12\sqrt{n}} \mathbb{E} \left(\sum_{i=1}^d \left| \sum_{l=1}^d \left[[I(\theta_0)]^{-\frac{1}{2}} \right]_{il} \left(\frac{\partial}{\partial \theta_l} \log f(X'_1 | \theta_0) - \frac{\partial}{\partial \theta_l} \log f(X_1 | \theta_0) \right) \right| \right)^3 \\
&= \frac{\|h\|_3}{12\sqrt{n}} \mathbb{E} \left(\left| \frac{X'_1 - \mu}{\sigma} - \frac{X_1 - \mu}{\sigma} \right| + \frac{1}{\sqrt{2}\sigma^2} |(X'_1 - \mu)^2 - (X_1 - \mu)^2| \right)^3 \\
&\leq \frac{\|h\|_3}{3\sqrt{n}} \left[\mathbb{E} \left| \frac{X'_1 - X_1}{\sigma} \right|^3 + \mathbb{E} \left| \frac{(X'_1 - \mu)^2 - (X_1 - \mu)^2}{\sqrt{2}\sigma^2} \right|^3 \right] \\
&\leq \frac{8\|h\|_3}{3\sqrt{n}} \left[\frac{\mathbb{E}|X_1|^3}{\sigma^3} + \frac{\mathbb{E}(X_1 - \mu)^6}{2\sqrt{2}\sigma^6} \right] = \frac{8\|h\|_3}{3\sqrt{n}} \left[\frac{2\sqrt{2}}{\sqrt{\pi}} + \frac{15}{2\sqrt{2}} \right] \leq \frac{19}{\sqrt{n}} \|h\|_3. \quad (82)
\end{aligned}$$

For the choice of $\epsilon = \epsilon_0$ as in (R.C.3), (70), (71) and (72) require that $0 < \epsilon < \sigma^2$. Due to a trade-off between the first and the last term of the bound in (6), we choose $\epsilon = \frac{\sigma^2}{2}$. Then, for the last term of (6), using (68) we obtain that

$$\frac{2\|h\|}{\epsilon^2} \left(\sum_{j=1}^d \mathbb{E} \left[\left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0j} \right)^2 \right] \right) = \frac{8\|h\|}{\sigma^4} \left(\frac{\sigma^2}{n} + \frac{\sigma^4}{n} \left(2 - \frac{1}{n} \right) \right) \leq \frac{8\|h\|}{n\sigma^2} (1 + 2\sigma^2). \quad (83)$$

Using the results in (69), (79), (80), (81), (82) and (83) we get the result of the corollary. ■

Proof of Corollary 3.1.

Part a). The probability density function is

$$f(x|\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

with $\alpha, \beta > 0$ and $x \in [0, 1]$. Hence, for $j, k \in \mathbb{Z}^+$

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \log f(x|\theta) &= \Psi(\alpha + \beta) - \Psi(\alpha) + \log(x), & \frac{\partial}{\partial \beta} \log f(x|\theta) &= \Psi(\alpha + \beta) - \Psi(\beta) + \log(1-x) \\
\frac{\partial^{j+1}}{\partial \alpha^{j+1}} \log f(x|\theta) &= \Psi_j(\alpha + \beta) - \Psi_j(\alpha), & \frac{\partial^{j+1}}{\partial \beta^{j+1}} \log f(x|\theta) &= \Psi_j(\alpha + \beta) - \Psi_j(\beta) \\
\frac{\partial^{k+j}}{\partial \alpha^k \partial \beta^j} \log f(x|\theta) &= \Psi_{k+j-1}(\alpha + \beta). \quad (84)
\end{aligned}$$

The assumptions (Con.1)-(Con.3) are satisfied. For (Con.1) notice that $x \in [0, 1]$. For (Con.2), let $\epsilon_0 = \epsilon = \epsilon(\boldsymbol{\theta}_0) : 0 < \epsilon < \min\{\alpha, \beta\}$. In addition, $\sum_{i=1}^{\infty} \frac{1}{i^4} = \frac{\pi^4}{90}$ and $\sum_{i=1}^{\infty} \frac{1}{i^3} = \zeta(3)$, where $\zeta(\cdot)$ is the Riemann zeta function and $\zeta(3)$ is known as Apéry's constant, which is an irrational number (≈ 1.202). For the calculations that follow, $\zeta(3) < 1.21$ is used. With $\boldsymbol{\theta} = (\theta_1, \theta_2)$, using (33), yields

$$\begin{aligned} \sup_{\substack{\boldsymbol{\theta}: |\theta_1 - \alpha| < \epsilon \\ |\theta_2 - \beta| < \epsilon}} \left| \frac{\partial^3}{\partial \theta_1^3} \log f(x|\boldsymbol{\theta}) \right| &\leq \beta |\Psi_3(\alpha - \epsilon)| = 6\beta \sum_{k=0}^{\infty} \frac{1}{(\alpha - \epsilon + k)^4} \leq 6\beta \left[\frac{1}{(\alpha - \epsilon)^4} + \sum_{k=1}^{\infty} \frac{1}{k^4} \right] \\ &\leq \frac{6\beta}{(\alpha - \epsilon)^4} + \frac{\beta\pi^4}{15} := M_{111} \end{aligned} \quad (85)$$

Following a similar process,

$$\begin{aligned} \sup_{\substack{\boldsymbol{\theta}: |\theta_1 - \alpha| < \epsilon \\ |\theta_2 - \beta| < \epsilon}} \left| \frac{\partial^3}{\partial \theta_2^3} \log f(x|\boldsymbol{\theta}) \right| &\leq \frac{6\alpha}{(\beta - \epsilon)^4} + \frac{\alpha\pi^4}{15} := M_{222} \\ \sup_{\substack{\boldsymbol{\theta}: |\theta_1 - \alpha| < \epsilon \\ |\theta_2 - \beta| < \epsilon}} \left| \frac{\partial^3}{\partial \theta_1^2 \partial \theta_2} \log f(x|\boldsymbol{\theta}) \right| &= \sup_{\substack{\boldsymbol{\theta}: |\theta_1 - \alpha| < \epsilon \\ |\theta_2 - \beta| < \epsilon}} \left| \frac{\partial^3}{\partial \theta_1 \partial \theta_2^2} \log f(x|\boldsymbol{\theta}) \right| \\ &\leq \frac{2}{(\alpha + \beta - 2\epsilon)^3} + 2\zeta(3) < \frac{2}{(\alpha + \beta - 2\epsilon)^3} + 2.42 \\ &:= M_{112} = M_{121} = M_{211} = M_{122} = M_{212} = M_{221}. \end{aligned} \quad (86)$$

Therefore, (Con.2) is also satisfied. Using (84), the expected Fisher Information matrix is

$$I(\boldsymbol{\theta}_0) = \begin{pmatrix} \Psi_1(\alpha) - \Psi_1(\alpha + \beta) & -\Psi_1(\alpha + \beta) \\ -\Psi_1(\alpha + \beta) & \Psi_1(\beta) - \Psi_1(\alpha + \beta) \end{pmatrix}.$$

The inverse of this 2×2 matrix is

$$[I(\boldsymbol{\theta}_0)]^{-1} = \frac{1}{\delta_I} \begin{pmatrix} \Psi_1(\beta) - \Psi_1(\alpha + \beta) & \Psi_1(\alpha + \beta) \\ \Psi_1(\alpha + \beta) & \Psi_1(\alpha) - \Psi_1(\alpha + \beta) \end{pmatrix}$$

and

$$[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} = \frac{1}{[\delta_I (C_2(\alpha, \beta) + C_2(\beta, \alpha))]^{\frac{1}{2}}} \begin{pmatrix} C_2(\beta, \alpha) & \Psi_1(\alpha + \beta) \\ \Psi_1(\alpha + \beta) & C_2(\alpha, \beta) \end{pmatrix}. \quad (87)$$

For $M = M_B$, as defined in (Con.3), we have that

$$\begin{aligned} M_B &= \frac{1}{\delta_I} \sup \{ \Psi_1(\alpha + \beta), \Psi_1(\alpha) - \Psi_1(\alpha + \beta), \Psi_1(\beta) - \Psi_1(\alpha + \beta) \} \\ &= \frac{1}{\delta_I} \sup \{ \Psi_1(\alpha + \beta), \Psi_1(\min\{\alpha, \beta\}) - \Psi_1(\alpha + \beta) \}, \end{aligned}$$

as $\Psi_1(\cdot)$ is a decreasing function. Having found M_B and $[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}$ and using (85) and (86) with the notation (34), (Con.3) is satisfied for

$$\begin{aligned} n &\geq \frac{4}{m^2} \left(\frac{mM_B (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta)}{2\sqrt{\delta_I} (C_2(\alpha, \beta) + C_2(\beta, \alpha))} \right. \\ &\quad \left. + \left[\left(\frac{mM_B (\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta)}{2\sqrt{\delta_I} (C_2(\alpha, \beta) + C_2(\beta, \alpha))} \right)^2 + 8M_B \right]^{\frac{1}{2}} \right)^2. \end{aligned}$$

We proceed with the bound for $E \left[\sum_{j=1}^2 \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right]$. Firstly, as the second-order partial derivatives of the log-likelihood function for the Beta distribution are not random,

$$\text{Var} \left(\frac{\partial^2}{\partial \alpha^2} \log f(X|\boldsymbol{\theta}_0) \right) = \text{Var} \left(\frac{\partial^2}{\partial \beta^2} \log f(X|\boldsymbol{\theta}_0) \right) = \text{Var} \left(\frac{\partial^2}{\partial \alpha \partial \beta} \log f(X|\boldsymbol{\theta}_0) \right) = 0$$

and hence $v = 0$. Therefore, for the Beta distribution

$$E \left(\sum_{j=1}^2 \left(\hat{\theta}_n(\mathbf{X})_j - \theta_{0,j} \right)^2 \right) \leq U_1 = \frac{\sqrt{4\omega\gamma}}{2\sqrt{n\omega}} = \sqrt{\frac{\gamma}{n\omega}}.$$

Our focus now turns to find an upper bound for γ . The first quantity for γ in (28) is

$$\sum_{j=1}^d \left| \left[[I(\boldsymbol{\theta}_0)]^{-1} \right]_{jj} \right| = \frac{1}{\delta_I} [\Psi_1(\beta) + \Psi_1(\alpha) - 2\Psi_1(\alpha + \beta)]. \quad (88)$$

For the second term, notice that from (5.12) of Anastasiou and Reinert (2015),

$$E \left[\left(\frac{\partial}{\partial \alpha} \log f(X_1|\boldsymbol{\theta}_0) \right)^4 \right] \leq 8C_1(\alpha, \beta), \quad E \left[\left(\frac{\partial}{\partial \beta} \log f(X_1|\boldsymbol{\theta}_0) \right)^4 \right] \leq 8C_1(\beta, \alpha), \quad (89)$$

with $C_1(x, y)$ as in (34). Using (89) and (87) gives after calculations

$$\begin{aligned} & \left[\text{Var} \left(\left(\sum_{k=1}^2 \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{1k} \frac{\partial}{\partial \theta_k} \log f(X_1|\boldsymbol{\theta}_0) \right)^2 \right) \right]^{\frac{1}{2}} \\ & + \left[\text{Var} \left(\left(\sum_{k=1}^2 \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{2k} \frac{\partial}{\partial \theta_k} \log f(X_1|\boldsymbol{\theta}_0) \right)^2 \right) \right]^{\frac{1}{2}} \\ & \leq \frac{8}{\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ \left[(C_2(\beta, \alpha))^4 C_1(\alpha, \beta) + (\Psi_1(\alpha + \beta))^4 C_1(\beta, \alpha) \right]^{\frac{1}{2}} \right. \\ & \quad \left. + \left[(\Psi_1(\alpha + \beta))^4 C_1(\alpha, \beta) + (C_2(\alpha, \beta))^4 C_1(\beta, \alpha) \right]^{\frac{1}{2}} \right\}. \quad (90) \end{aligned}$$

For the third term of the expression for γ as in (28) since $d = 2$,

$$\begin{aligned} & \left[\text{Var} \left(\sum_{q=1}^2 \sum_{v=1}^2 \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{2q} \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{1v} \frac{\partial}{\partial \theta_q} \log f(X_1|\boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_v} \log f(X_1|\boldsymbol{\theta}_0) \right) \right]^{\frac{1}{2}} \\ & \leq \left[E \left(\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{12} \left\{ \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{11} \left(\frac{\partial}{\partial \alpha} \log f(X_1|\boldsymbol{\theta}_0) \right)^2 + \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{22} \left(\frac{\partial}{\partial \beta} \log f(X_1|\boldsymbol{\theta}_0) \right)^2 \right\} \right. \right. \\ & \quad \left. \left. + \frac{\partial}{\partial \alpha} \log f(X_1|\boldsymbol{\theta}_0) \frac{\partial}{\partial \beta} \log f(X_1|\boldsymbol{\theta}_0) \left[\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{22} \left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{11} + \left(\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} \right]_{12} \right)^2 \right] \right)^2 \right]^{\frac{1}{2}}. \quad (91) \end{aligned}$$

Using now the known inequality, $(\sum_{i=1}^n a_i)^2 \leq n \sum_{i=1}^n a_i^2$ for $a_i \in R$, and also the Cauchy-Schwarz inequality yields

$$(91) \leq \frac{\sqrt{24}}{\delta_I [C_2(\alpha, \beta) + C_2(\beta, \alpha)]} \left\{ (\Psi_1(\alpha + \beta))^2 [C_3(\alpha, \beta) + C_3(\beta, \alpha)] \right. \\ \left. + 2 [(C_2(\alpha, \beta))^2 (C_2(\beta, \alpha))^2 + (\Psi_1(\alpha + \beta))^4] \sqrt{C_1(\alpha, \beta) C_1(\beta, \alpha)} \right\}^{\frac{1}{2}}. \quad (92)$$

The inequalities in (88), (90) and (92) show that $\gamma \leq \gamma_B$ as in (34). To calculate ω ,

$$\sum_{l=1}^2 \sum_{k=1}^2 \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{lk} \left| \sum_{m=1}^2 \sum_{i=1}^2 M_{kim} \right| = \sum_{k=1}^2 \left[\left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{1k} + \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{2k} \right] \sum_{m=1}^2 \sum_{i=1}^2 M_{kim} \\ = \frac{(\Psi_1(\beta) + \sqrt{\delta_I}) C_4(\beta, \alpha) + (\Psi_1(\alpha) + \sqrt{\delta_I}) C_4(\alpha, \beta)}{\sqrt{\delta_I} (C_2(\alpha, \beta) + C_2(\beta, \alpha))} \quad (93)$$

shows that $\omega = \omega_B$ as defined in (34). Since $\omega_B > 0$ for n satisfying (Con.3), we conclude that

$$\mathbb{E} \left(\sum_{j=1}^2 (\hat{\theta}_n(X)_j - \theta_{0j})^2 \right) \leq \frac{\sqrt{4\omega_B \gamma_B}}{2\sqrt{n\omega_B}} = \frac{\sqrt{\omega_B \gamma_B}}{\sqrt{n\omega_B}} = \sqrt{\frac{\gamma_B}{n\omega_B}}. \quad (94)$$

Part **b**). For (36) we use the general expression of the bound in (32). Upper bounds for the first two terms of (32) have already been obtained in (90) and (92). Now, for ease of presentation let

$$K_1(X_1, X'_1) := \frac{\partial}{\partial \alpha} \log f(X'_1 | \theta_0) - \frac{\partial}{\partial \alpha} \log f(X_1 | \theta_0) \\ K_2(X_1, X'_1) := \frac{\partial}{\partial \beta} \log f(X'_1 | \theta_0) - \frac{\partial}{\partial \beta} \log f(X_1 | \theta_0). \quad (95)$$

Using (95) and the inequality in (24), we have for the third term in (32),

$$\mathbb{E} \left(\sum_{i=1}^2 \left| \sum_{l=1}^2 [I(\theta_0)]^{-\frac{1}{2}} \right|_{il} \left(\frac{\partial}{\partial \theta_l} \log f(X'_1 | \theta_0) - \frac{\partial}{\partial \theta_l} \log f(X_1 | \theta_0) \right) \right)^3 \\ = 4\mathbb{E} \left(\left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{11} (K_1(X_1, X'_1)) + [I(\theta_0)]^{-\frac{1}{2}}_{12} (K_2(X_1, X'_1)) \right|^3 \\ + \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{22} (K_1(X_1, X'_1)) + [I(\theta_0)]^{-\frac{1}{2}}_{21} (K_2(X_1, X'_1)) \right|^3 \\ \leq 16 \left(\left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{11}^3 + \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{12}^3 \right) \mathbb{E} (|K_1(X_1, X'_1)|^3) \\ + \left(\left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{12}^3 + \left| [I(\theta_0)]^{-\frac{1}{2}} \right|_{22}^3 \right) \mathbb{E} (|K_2(X_1, X'_1)|^3)$$

$$\leq \frac{128}{[\delta_I(C_2(\alpha, \beta) + C_2(\beta, \alpha))]^{\frac{3}{2}}} \left\{ 8^{\frac{3}{4}} [C_1(\alpha, \beta)]^{\frac{3}{4}} \left([C_2(\beta, \alpha)]^3 + [\Psi_1(\alpha + \beta)]^3 \right) + 8^{\frac{3}{4}} [C_1(\beta, \alpha)]^{\frac{3}{4}} \left([C_2(\alpha, \beta)]^3 + [\Psi_1(\alpha + \beta)]^3 \right) \right\}. \quad (96)$$

The above inequalities are a result of (24). To be more specific, for the last inequality in (96) we use that

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_j} \log f(X'_1 | \boldsymbol{\theta}_0) - \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right|^3 &\leq \left(\left| \frac{\partial}{\partial \theta_j} \log f(X'_1 | \boldsymbol{\theta}_0) \right| + \left| \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right| \right)^3 \\ &\leq 4 \left(\left| \frac{\partial}{\partial \theta_j} \log f(X'_1 | \boldsymbol{\theta}_0) \right|^3 + \left| \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right|^3 \right) \end{aligned}$$

and therefore,

$$\begin{aligned} \mathbb{E} \left(\left| \frac{\partial}{\partial \theta_j} \log f(X'_1 | \boldsymbol{\theta}_0) - \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right|^3 \right) &\leq 4 \mathbb{E} \left(\left| \frac{\partial}{\partial \theta_j} \log f(X'_1 | \boldsymbol{\theta}_0) \right|^3 + \left| \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right|^3 \right) \\ &= 8 \mathbb{E} \left(\left| \frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right|^3 \right) \leq 8 \left[\mathbb{E} \left(\frac{\partial}{\partial \theta_j} \log f(X_1 | \boldsymbol{\theta}_0) \right)^4 \right]^{\frac{3}{4}}. \end{aligned}$$

Thus, applying the results of (90), (92), (93), (94) and (96) to (32) Corollary 3.1 follows. \blacksquare

References

- Anastasiou, A. and G. Reinert (2015). Bounds for the normal approximation of the maximum likelihood estimator. *ArXiv 1411.2391*.
- Billingsley, P. (1961). Statistical Methods in Markov Chains. *The Annals of Mathematical Statistics* **32**, No.1, 12–40.
- Casella, G. and R. L. Berger (2002). *Statistical Inference* (Second ed.). Brooks/Cole, Cengage Learning, Duxbury, Pacific Grove.
- Davison, A. C. (2008). *Statistical Models* (First ed.). Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* **13**, No.1, 342–368.
- Hoadley, B. (1971). Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. *The Annals of Mathematical Statistics* **42**, No.6, 1977–1991.
- Makelainen, T., K. Schmidt, and G. P. H. Styan (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *The Annals of Statistics* **9**, No.4, 758–767.
- Reinert, G. and A. Röllin (2009). Multivariate normal approximation with Stein’s method of exchangeable pairs under a general linearity condition. *The Annals of Probability* **37**, No.6, 2150–2173.